**FACULDADE SENAI GASPAR RICARDO JÚNIOR**

BIANCA NUNES

JÚLIA HARUMI NASCIMENTO

SARA FERREIRA ROSMANINHO ROZA

**FARMA MAIS**

Projeto: Ciências de Dados

Sorocaba – SP

2025

**SUMÁRIO**

[**1. Introdução 3**](#_hseu2fixu52d)

[1.1. Séries Temporais 3](#_ifnzw7n5fkjw)

[1.2. Visualizações com gráficos 3](#_cbzgj51of3n)

[1.3. Medidas de Dispersão 3](#_xdagfc8hfyi5)

[1.4. Correlação entre variáveis 4](#_s481tg51fr1d)

[1.5. Tipos de amostragem 4](#_v90w6djt045v)

[1.6. Medidas de tendência central 4](#_hobwsdgxx9uv)

[**2. Referencial Teórico 4**](#_blwuf76zdiff)

[2.1. Conceitos Estatísticos Essenciais 5](#_7womnsjdrfwt)

[2.2. Conceitos Computacionais e Visualização 6](#_gxbxinvbdz6s)

[2.3. Bibliotecas Python Aplicadas 6](#_a9ks8aa302pq)

[**3. Metodologia 7**](#_5ps86beagey2)

[3.1. Descrição da base de dados 7](#_bv3057ghx4v1)

[3.2. Etapas de Tratamento e Limpeza de Dados 8](#_afqmy656i8nw)

[3.3. Ferramentas e Processos Utilizados 8](#_j72aj7orv7s6)

[**4. Análise de Dados 9**](#_gogi2ijnbp32)

[4.1. Séries Temporais 10](#_oipkpdoy1qav)

[4.2. Visualizações com Gráficos Estatísticos 11](#_z8lsbw1keaxg)

[4.2.1. Visualização com Gráficos de Barra 12](#_bmx2tnp3sz1b)

[4.2.2. Visualização com Gráficos BoxPlot 14](#_f71dbwt7f7zu)

[4.2.3. Visualização com Histograma 15](#_572mjr3m92nq)

[4.3. Medidas de Tendência Central (Média, Mediana e Moda) 15](#_hghvlze6uzm7)

[4.4. Medidas de Dispersão 16](#_np9gvo8ui7c6)

[4.4.1. Amplitude 16](#_m9xs6v82bv9)

[4.4.2. Variância 17](#_gzx8td7c27xg)

[4.4.3. Desvio Padrão 18](#_q2xwubj1qj5b)

[4.5. Correlação entre Variáveis 18](#_4p31jol8nz9i)

[4.6. Tipos de Amostragem 20](#_uibt7cgikl9g)

[**5. Conclusão 21**](#_1xfljbhlofy6)

[5.1. Limitações 21](#_jf5a0dkyg64k)

[5.2. Sugestões para Trabalhos Futuros 22](#_s22hhlzcyaoz)

[**6. Apêndices 22**](#_4nw8a1k1kmbd)

[**7. Referências 23**](#_8wq1fh2859po)

# Introdução

Para esse projeto o estabelecimento escolhido para ter os dados (fictícios) analisados é o de farmácia e o objetivo das análises que serão apresentadas para esse segmento é extrair insights valiosos e tomar decisões estratégicas mais embasadas a partir dos dados disponíveis. Ao aplicar as técnicas de ciência de dados que serão listadas logo a seguir em seis tópicos, a intenção é compreender o comportamento dos clientes, o desempenho dos produtos e as dinâmicas do mercado farmacêutico.

## 1.1. Séries Temporais

A análise de dados temporais para identificar padrões, sazonalidades ou tendências. Permite que sejam analisados dados de vendas de medicamentos ao longo do tempo (dias, semanas, meses, anos), identifique picos de demanda, sazonalidades (por exemplo, mais vendas de antigripais no inverno) e tendências de crescimento ou declínio em determinados produtos, com isso, é possível otimizar o estoque e planejar promoções.

## 1.2. Visualizações com gráficos

Visualizações com Gráficos: Estatísticos Criação de histogramas, boxplots, gráficos de dispersão podem ser usados para entender e comunicar os padrões identificados pelas séries temporais e outras análises, os gráficos são indispensáveis, utilizando eles é mais fácil identificar a evolução das vendas, histogramas para visualizar a idade dos clientes em relação a frequência de compras, ou gráficos de barras para comparar a demanda em diferentes meses. A visualização eficaz torna os dados compreensíveis e facilita a tomada de decisões.

## 1.3. Medidas de Dispersão

Medidas de Dispersão: Análise de variabilidade dos dados utilizando amplitude, variância e desvio padrão. Enquanto as medidas de tendência central mostram o "típico", as medidas de dispersão revelam o quão variáveis são os dados, sendo possível entender a variabilidade nas vendas de um medicamento, por exemplo, é crucial para evitar falta ou excesso de estoque.

## 1.4. Correlação entre variáveis

Correlação entre Variáveis: Avaliação da relação entre variáveis com coeficientes de Pearson ou Spearman. Este tópico permite que descubra se existe uma relação entre diferentes fatores no gerenciamento. Por exemplo, investigar se há uma correlação entre o número de vendas de um determinado medicamento e o mês do ano, isso pode revelar insights para otimizar o atendimento ou a organização da equipe.

## 1.5. Tipos de amostragem

Tipos de Amostragem: Aplicação de técnicas como amostragem aleatória, sistemática e estratificada, essa aplicação de amostragem é fundamental quando não é possível analisar todos os dados. O objetivo é garantir que a amostra coletada seja representativa da população total para análises mais aprofundadas, sem precisar processar todos os dados brutos, economizando tempo e recursos.

## 1.6. Medidas de tendência central

Cálculo e interpretação de média, mediana e moda se encaixa nesse projeto por serem medidas simples, mas poderosas, ajudam a entender o comportamento típico dos dados, podem calcular a média de vendas diárias de um medicamento, a mediana do número de clientes por hora, ou a moda do horário de pico de vendas.

# Referencial Teórico

Por meio da análise de grandes volumes de dados, é possível extrair informações valiosas que auxiliam na tomada de decisões estratégicas. Na área farmacêutica, a aplicação de técnicas de ciência de dados permite, por exemplo, o monitoramento do estoque, a análise do comportamento de compra dos clientes, a previsão de demanda e a identificação de produtos com maior rotatividade ou sazonalidade.

## 2.1. Conceitos Estatísticos Essenciais

Para transformar dados brutos em insights acionáveis, utilizamos diversas ferramentas da estatística. Primeiramente, as medidas de tendência central indicam o “centro” dos dados e são essenciais para resumir conjuntos numéricos. A média, por exemplo, é calculada pela soma de todos os valores dividida pelo total de itens, como a média de vendas diárias, sendo sensível a valores extremos. Já a mediana corresponde ao valor central em dados ordenados, como a idade mediana dos clientes, e é robusta a valores extremos. A moda representa o valor mais frequente em um conjunto, como o produto mais vendido.

Além disso, as medidas de dispersão quantificam a variabilidade dos dados. A amplitude, que é a diferença entre o maior e o menor valor, fornece uma noção simples dessa variação. A variância e o desvio padrão medem o quanto os dados se afastam da média, sendo que um alto desvio padrão nas vendas indica grande oscilação.

Outro aspecto importante é a correlação entre variáveis, que avalia a relação entre dois conjuntos de dados. O coeficiente de Pearson é utilizado para analisar relações lineares, como quando as vendas de protetor solar e pós-sol aumentam simultaneamente. Já o coeficiente de Spearman é indicado para relações não lineares ou para variáveis ordinais.

Na análise temporal, as séries temporais permitem o estudo dos dados ao longo do tempo para identificar padrões e sazonalidades, como o aumento nas vendas de antigripais durante o inverno ou o crescimento nas vendas de dermocosméticos. Por fim, os tipos de amostragem são fundamentais para selecionar subconjuntos representativos quando a população é muito grande. A amostragem aleatória simples garante que cada item tenha a mesma chance de ser selecionado, enquanto a amostragem estratificada divide a população em grupos homogêneos e amostra cada um deles, assegurando representatividade, por exemplo, por faixa etária dos clientes.

## 2.2. Conceitos Computacionais e Visualização

A manipulação e visualização dos dados são fundamentais para compreender e comunicar os resultados obtidos na análise. As visualizações por meio de gráficos estatísticos tornam os dados mais acessíveis e compreensíveis. Por exemplo, os histogramas são utilizados para mostrar a distribuição de uma variável numérica, como a idade dos clientes, enquanto os boxplots exibem a distribuição dos dados e permitem identificar outliers, como a variação da margem de lucro por categoria de produto. Já os gráficos de dispersão são eficazes para revelar a relação entre duas variáveis, como o valor total da venda em função da quantidade vendida. Para evidenciar tendências ao longo do tempo, os gráficos de linha são ideais, permitindo visualizar, por exemplo, as vendas totais por mês. Por fim, os gráficos de barras possibilitam a comparação de quantidades entre diferentes categorias, como a frequência de vendas por produto.

## 2.3. Bibliotecas Python Aplicadas

Essas bibliotecas foram as ferramentas essenciais para a análise, pois tornaram possível organizar, entender e visualizar os dados de forma prática, facilitando a descoberta de padrões e informações importantes para a farmácia.

Para a realização da análise de dados, foram utilizadas algumas bibliotecas fundamentais da linguagem Python. A biblioteca **Pandas** foi essencial para a manipulação e análise dos dados, permitindo trabalhar com estruturas como o *Data Frame*, que se assemelha a uma tabela, além de possibilitar carregar, limpar, transformar e agrupar os dados de maneira eficiente. A biblioteca **NumPy** serviu como base para operações numéricas, oferecendo suporte a vetores, matrizes e funções matemáticas de alto desempenho. Para a criação de visualizações, foi utilizada a biblioteca **Matplotlib.pyplot**, responsável pela geração de gráficos estáticos personalizados. De forma complementar, a biblioteca **Seaborn**, construída sobre a Matplotlib, proporcionou uma interface mais simplificada e estética para a criação de gráficos estatísticos mais sofisticados e informativos. Além disso, o módulo **Statistics**, nativo do Python, foi empregado para a realização de cálculos estatísticos básicos, como média, mediana e moda, que auxiliaram na análise descritiva dos dados.

Em resumo, a combinação dessas técnicas estatísticas com o poder das bibliotecas Python permitiu à farmácia analisar profundamente suas operações, desde o comportamento do cliente e a gestão de estoque até a lucratividade, auxiliando na tomada de decisões estratégicas baseadas em dados.

# Metodologia

Esta etapa apresenta a metodologia adotada no trabalho, descrevendo detalhadamente a base de dados utilizada, as etapas realizadas para o tratamento e limpeza dos dados, além das ferramentas e processos empregados para garantir a qualidade e a confiabilidade das informações analisadas.

## 3.1. Descrição da base de dados

A base de dados utilizada foi um conjunto de dados fictícios de vendas de farmácia, simulando registros transacionais. Ela foi criada programaticamente em Python para fins de demonstração e análise. O DataFrame gerado, denominado df\_farmacia\_excel ou df\_farmacia, contém como principais colunas aquelas que representam as informações essenciais para a análise.

O dataset aplicado na análise contém informações detalhadas sobre cada venda realizada, começando pelo identificador único da transação, denominado **ID\_Venda**. A data em que a venda foi efetuada, representada por **Data\_Venda**, possibilita a realização de análises temporais. Cada produto vendido está identificado pelo seu nome, como antialérgico ou protetor solar, registrado na variável **Produto**, e classificado em categorias, como dermocosméticos ou medicamentos tarjados, na variável **Categoria\_Produto**. O preço unitário de cada item é armazenado em **Preco\_Unitario**, enquanto a quantidade vendida em cada transação é registrada em **Quantidade\_Vendida**. O método de pagamento utilizado pelo cliente, como cartão de crédito ou Pix, consta na variável **Tipo\_Pagamento**. A idade do cliente no momento da compra está indicada em **Idade\_Cliente**, e a margem de lucro percentual associada ao produto é representada por **Margem\_Lucro\_Produto\_Perc**. O valor total da venda, calculado como o produto entre o preço unitário e a quantidade vendida, está registrado em **Valor\_Total\_Venda**. Para simular um cenário real de farmácia e testar a robustez das análises, a base de dados inclui anomalias introduzidas artificialmente nas variáveis de valor total da venda e quantidade vendida.

## 3.2. Etapas de Tratamento e Limpeza de Dados

Nesta metodologia, os processos foram relativamente simplificados, considerando a natureza sintética e controlada da base de dados. No entanto, os processos básicos incluem a geração, o tratamento e a análise dos dados.

Os dados foram inicialmente gerados em formato de dicionário Python e imediatamente convertidos para um DataFrame do Pandas, garantindo assim uma estrutura tabular organizada, essencial para as análises subsequentes. Em seguida, foi calculada a coluna Valor\_Total\_Venda, derivada a partir das variáveis Preco\_Unitario e Quantidade\_Vendida, um passo importante para obter métricas de negócio relevantes. Para simular a presença de outliers ou erros comuns em dados reais, algumas linhas do DataFrame tiveram seus valores de Valor\_Total\_Venda e Quantidade\_Vendida alterados intencionalmente para valores atipicamente altos, permitindo observar o impacto dessas anomalias nas análises, especialmente nas medidas de tendência central e dispersão. Posteriormente, o DataFrame foi exportado para um arquivo Excel (.xlsx), facilitando o uso dos dados em outras partes do script ou em diferentes ferramentas, simulando um fluxo de trabalho comum onde os dados podem ser provenientes de arquivos externos. Embora não tenha sido explicitamente realizada uma etapa de tratamento de valores ausentes, o uso de funções como .dropna() em cálculos de média, mediana e moda garantiu que apenas valores válidos fossem considerados, prevenindo erros e distorções nos resultados.

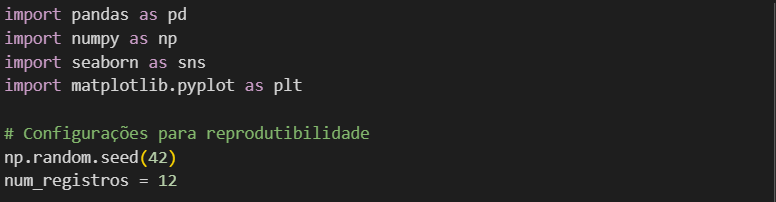
## 3.3. Ferramentas e Processos Utilizados

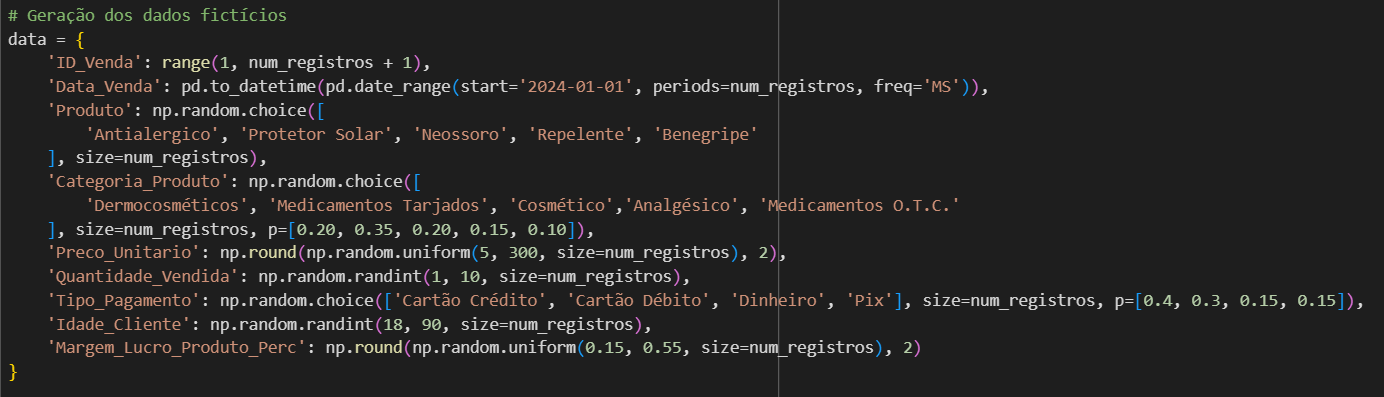
As análises foram realizadas utilizando a linguagem de programação Python, que oferece uma ampla variedade de recursos para manipulação e processamento de dados. Para complementar essa linguagem, foi utilizado um conjunto específico de bibliotecas especializadas, que permitiram desde a organização e limpeza dos dados até a criação de visualizações gráficas detalhadas, facilitando a interpretação dos resultados obtidos.

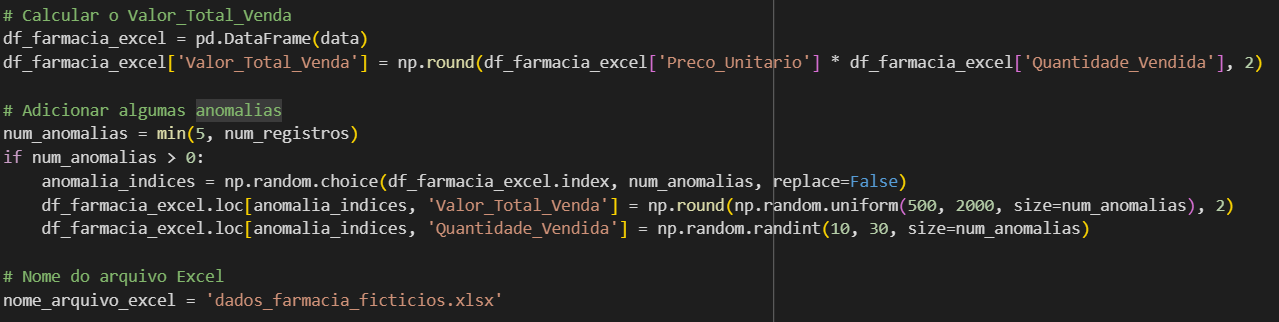
A biblioteca **Pandas** foi a ferramenta central para a manipulação e estruturação dos dados, sendo utilizada para criar o DataFrame, realizar agrupamentos por meio do método *groupby*, ordenar os dados com *sort\_values*, calcular novas colunas e exportar ou importar dados para arquivos Excel. Já o **NumPy** foi fundamental para operações numéricas de alto desempenho, empregado na geração de dados fictícios utilizando funções como *np.random.choice* e *np.random.uniform*, no arredondamento de valores com *np.round* e no cálculo de estatísticas básicas, como média e desvio padrão, por meio de *np.mean* e *np.std*. Para as visualizações gráficas, utilizou-se a biblioteca **Matplotlib.pyplot (plt)**, que controlou aspectos como o tamanho das figuras através de *plt.figure*, a adição de títulos com *plt.title*, a rotulagem dos eixos via *plt.xlabel* e *plt.ylabel*, a inserção de grades com *plt.grid* e o ajuste do layout com *plt.tight\_layout*. Complementarmente, a biblioteca **Seaborn (sns)**, construída sobre a Matplotlib, foi utilizada para criar gráficos estatísticos mais complexos e esteticamente agradáveis com menos código, sendo aplicada na geração de histogramas (*sns.histplot*), gráficos de linha para séries temporais (*sns.lineplot*), gráficos de barras (*sns.barplot*) e boxplots (*sns.boxplot*). Por fim, o módulo padrão do Python, **Statistics**, foi usado para calcular as medidas de tendência central, como média, mediana e moda, de forma direta para listas de valores.

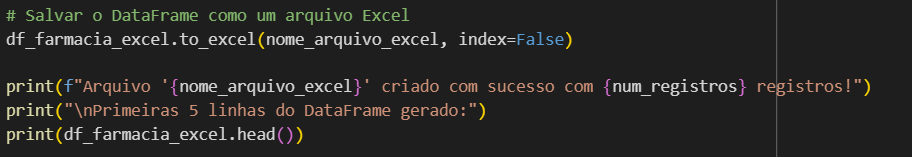
# Análise de Dados

Neste projeto, foram aplicadas seis ferramentas estatísticas e matemáticas, amplamente utilizadas na Ciência de Dados, com o objetivo de compreender melhor o comportamento de vendas, clientes e produtos em uma farmácia. Para isso, foi construído um DataFrame base, contendo valores fictícios simulados, que serviram como referência para a realização das análises e consultas subsequentes.



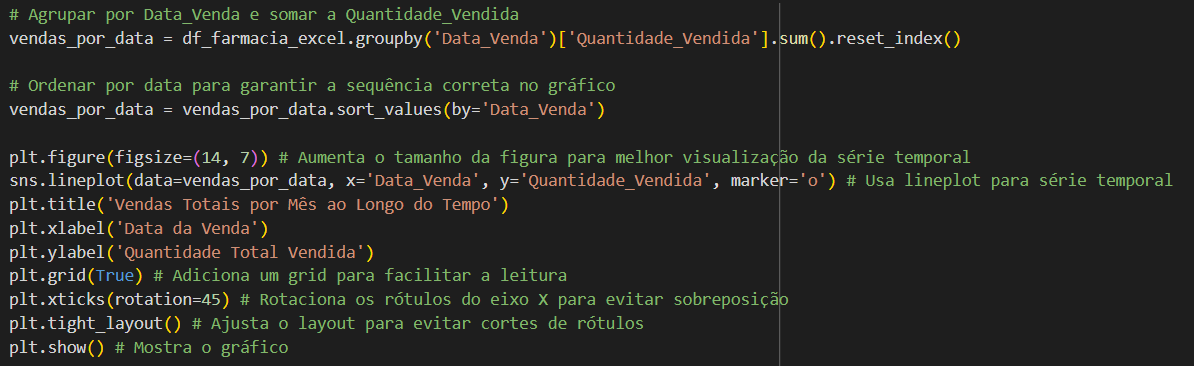


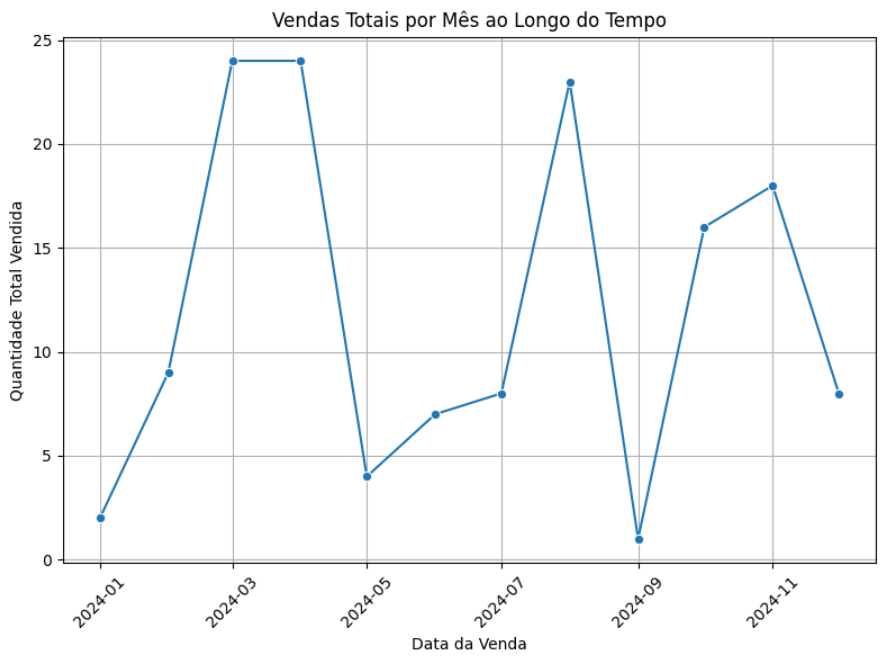
****

****

## 4.1. Séries Temporais

Os dados foram organizados pela coluna de data da venda para analisar como as vendas variaram ao longo do tempo. Com isso, conseguimos visualizar tendências e padrões, como possíveis aumentos de venda em certos períodos do ano. Essa análise ajuda a prever sazonalidades, como por exemplo um aumento de vendas de antigripais no inverno ou de protetores solares no verão.

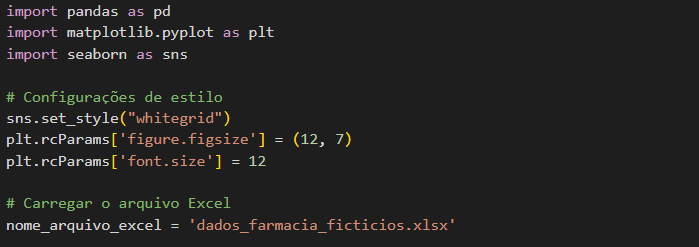


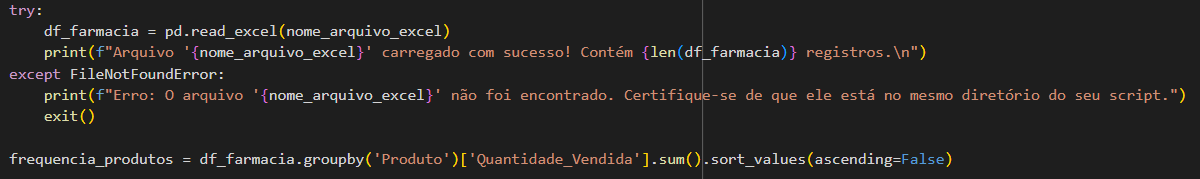


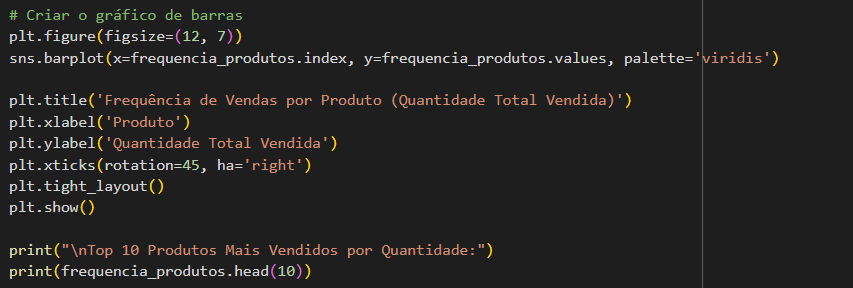
## 4.2. Visualizações com Gráficos Estatísticos

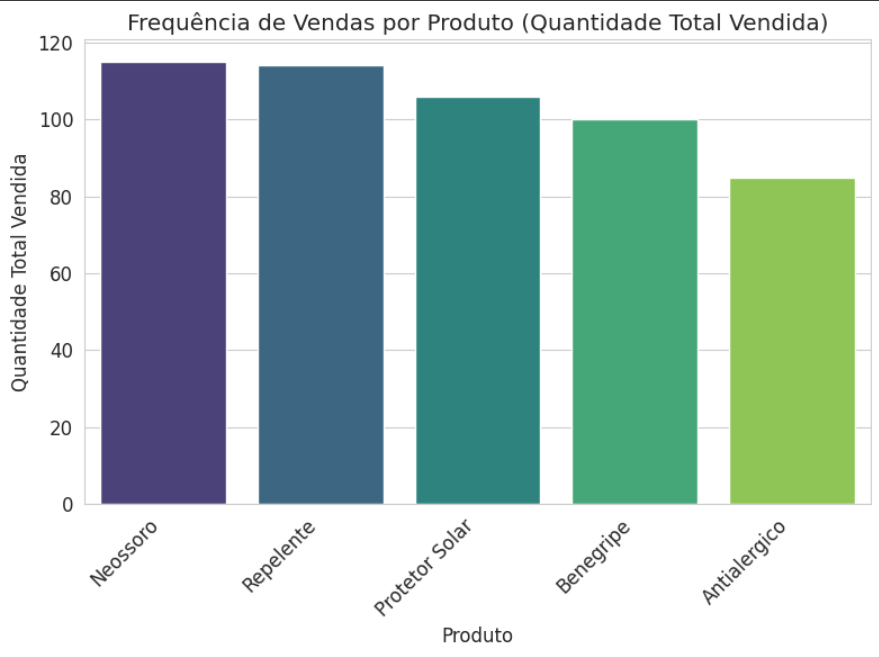
Utilizamos diferentes tipos de gráficos para facilitar a análise dos dados. Fizemos histogramas para ver a distribuição da idade dos clientes, boxplots para comparar variações de preço e lucro por categoria, e gráficos de dispersão para analisar a relação entre quantidade vendida e valor total. Esses gráficos tornaram os dados mais fáceis de entender e ajudaram a visualizar padrões que não aparecem só olhando os números.

### 4.2.1. Visualização com Gráficos de Barra



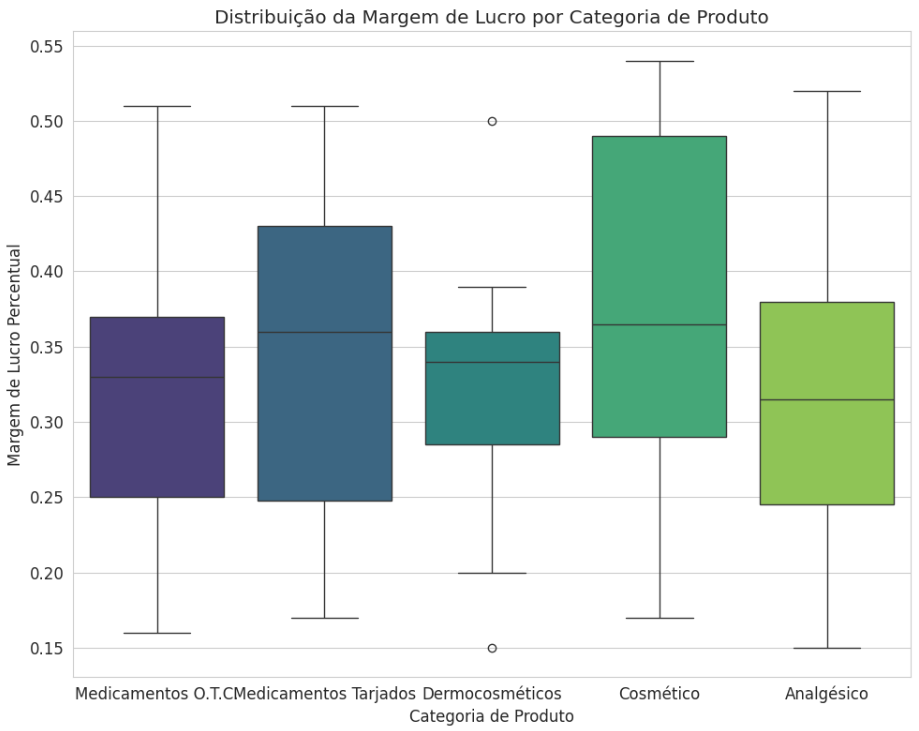






### 4.2.2. Visualização com Gráficos BoxPlot

### 

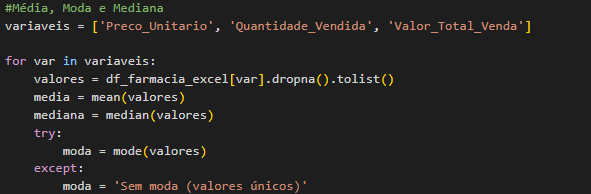


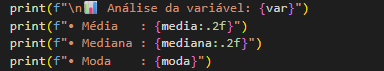
### 4.2.3. Visualização com Histograma

### 

## 4.3. Medidas de Tendência Central (Média, Mediana e Moda)

Com esses cálculos, conseguimos entender o que era mais comum ou central nos nossos dados. A média nos deu uma noção geral, a mediana ajudou a entender o valor típico mesmo com outliers, e a moda mostrou os valores mais frequentes, como a quantidade mais vendida ou o preço que mais apareceu. Foi muito útil para ter uma ideia geral do comportamento dos dados.

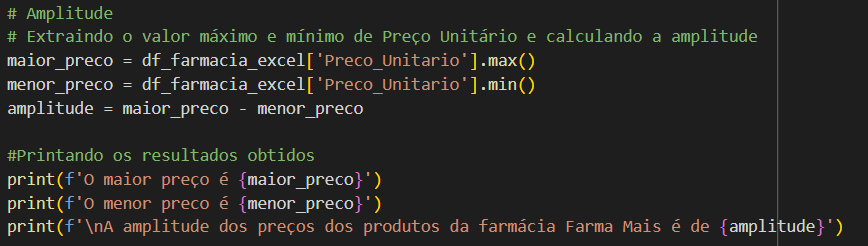




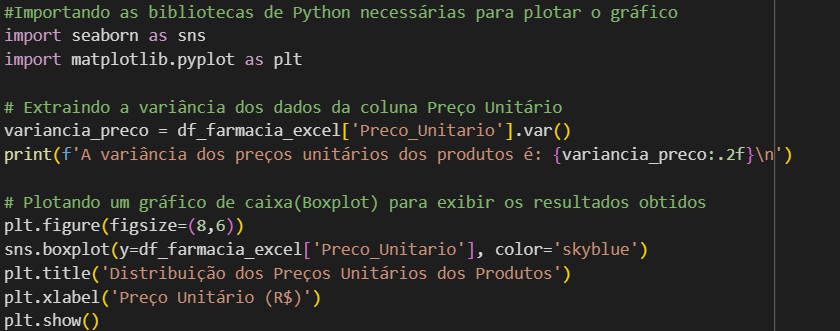
## 4.4. Medidas de Dispersão

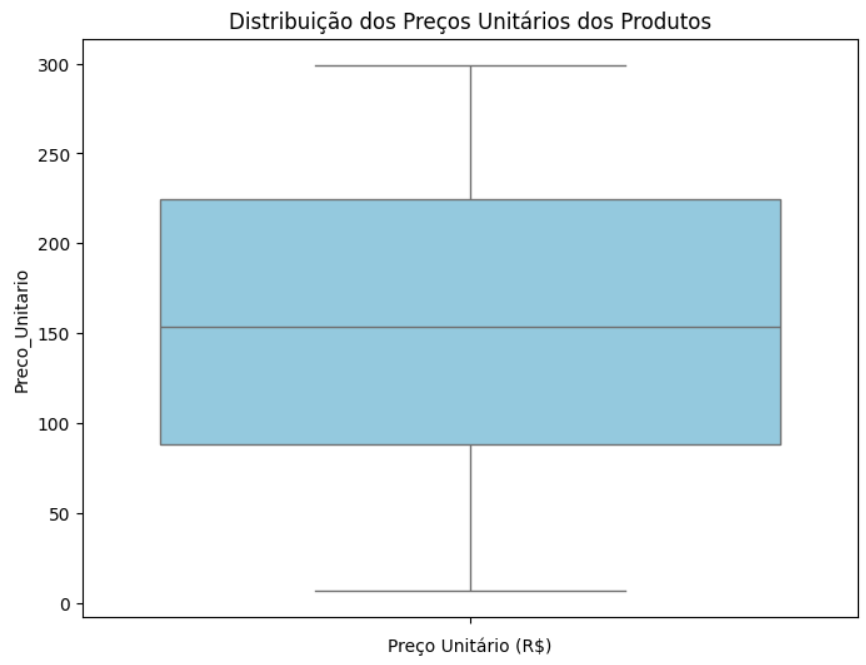
Analisamos a variabilidade nos dados usando o desvio padrão, a variância e a amplitude. Isso nos ajudou a entender o quanto os dados estavam concentrados ou espalhados. Por exemplo, ao calcular o desvio padrão do valor das vendas, percebemos se as vendas eram muito diferentes entre si ou se seguiam um padrão mais equilibrado.

### 4.4.1. Amplitude

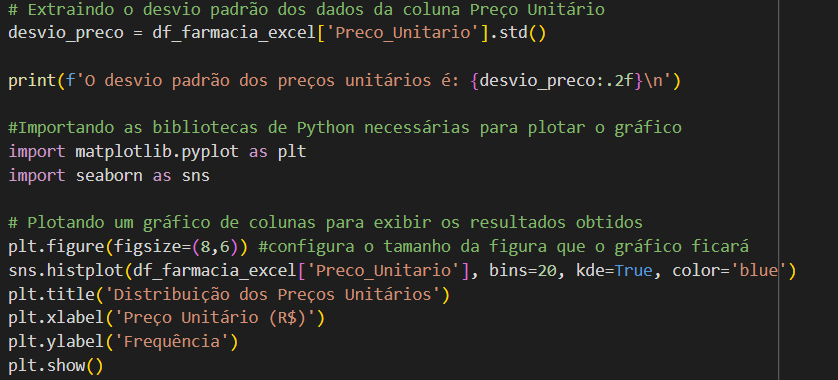


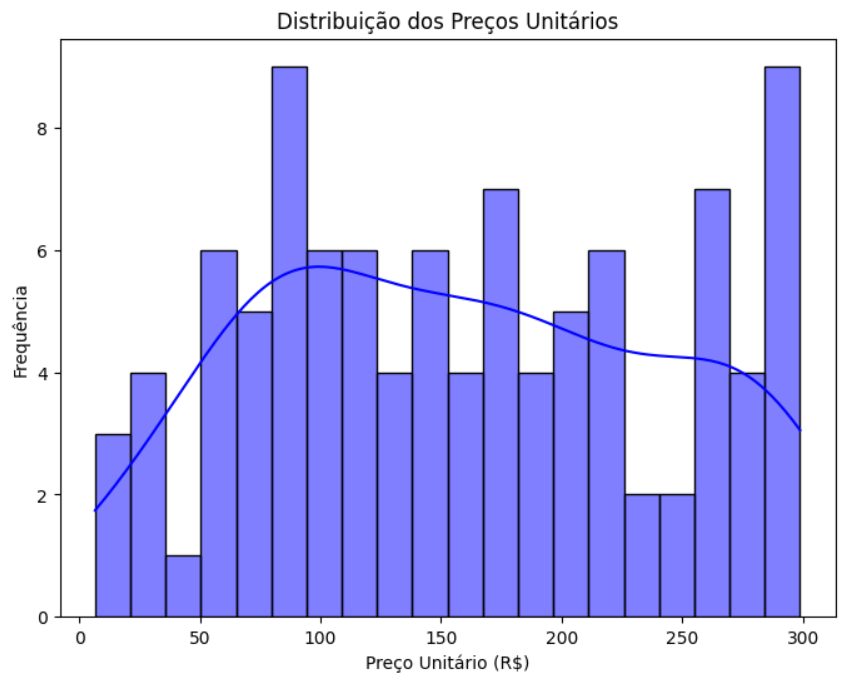
### 4.4.2. Variância





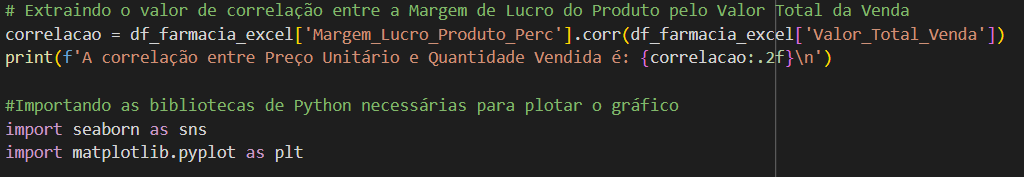
### 4.4.3. Desvio Padrão

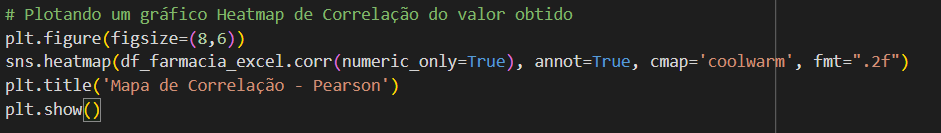


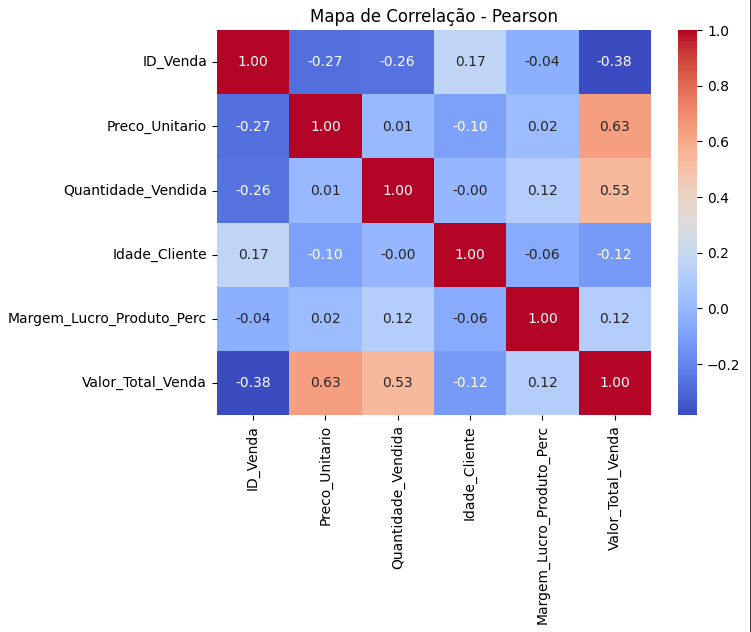


## 4.5. Correlação entre Variáveis

Verificamos se existiam relações entre diferentes variáveis, como entre a quantidade vendida e o valor total da venda. Usamos o coeficiente de Pearson para analisar essas relações. Descobrimos, por exemplo, que existe uma ligação direta entre vender mais unidades e ter um valor de venda maior, o que já era esperado, mas agora foi comprovado com os dados.

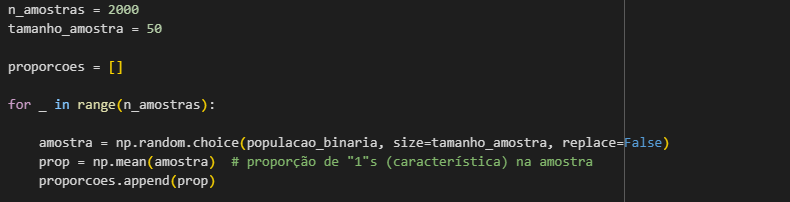
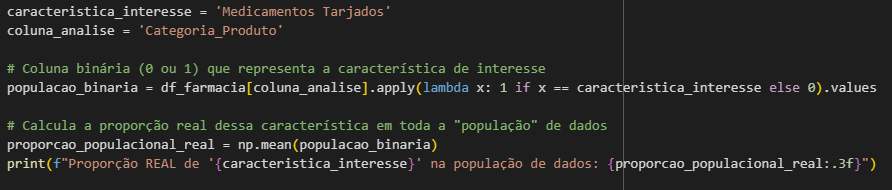


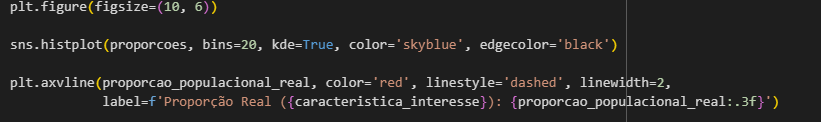


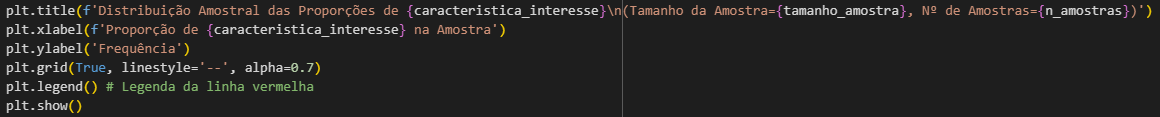


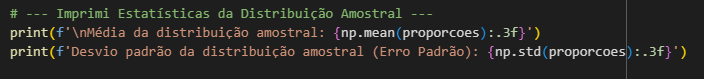
## 4.6. Tipos de Amostragem

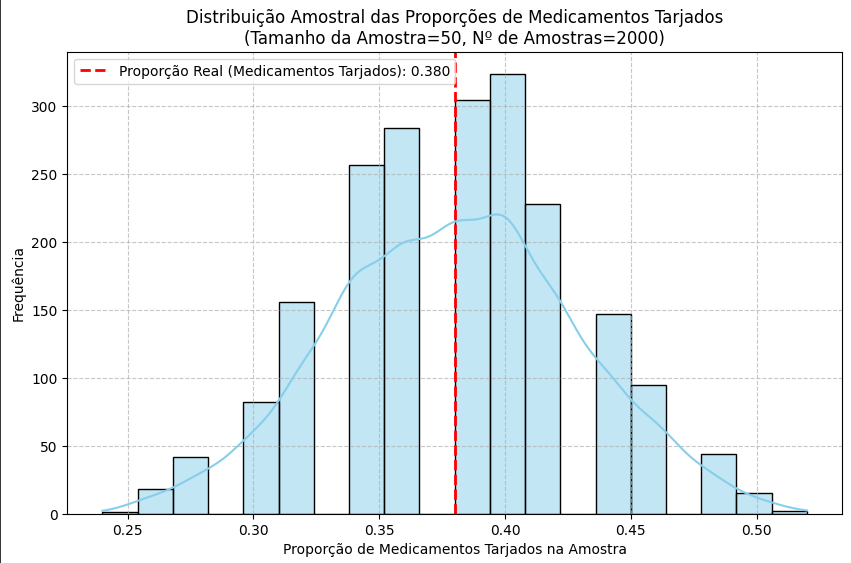
Mesmo usando dados fictícios, simulamos formas de fazer amostragem, como a aleatória simples e a estratificada (dividindo por faixa etária ou categoria de produto). Isso é importante em situações reais em que não dá para analisar todos os dados, e precisamos de uma amostra que represente bem o todo. A amostragem ajudaria muito a economizar tempo em análises maiores.











# Conclusão

Neste trabalho foram analisados dados simulados de uma farmácia, aplicando vários conceitos de estatística e programação em Python. Com a ajuda das bibliotecas Pandas, NumPy, Seaborn e Matplotlib, foi possível organizar os dados, fazer cálculos essenciais e montar gráficos que auxiliaram a entender os dados obtidos.

Foram calculadas medidas como a dispersão (que mostra o quanto os dados variam), correlações entre variáveis (para saber se um dado influencia no outro, como quantidade vendida e valor total), e também utilizamos as datas das vendas para organizar em formato de linha do tempo, o que facilitou a percepção de algumas tendências e padrões que se repetem.

Os gráficos ajudaram a mostrar de forma visual os resultados, como a distribuição de vendas, categorias de produtos mais vendidos, entre outros. Isso deixou a análise mais clara e fácil de entender.

## 5.1. Limitações

Por se tratar de um conjunto de dados simulados, alguns resultados não refletem, de forma precisa, a realidade operacional de uma farmácia. Além disso, a limitação de variáveis disponíveis e o tempo restrito para desenvolvimento impediram a realização de análises mais profundas e detalhadas.

## 5.2. Sugestões para Trabalhos Futuros

Caso este projeto venha a ser expandido, algumas melhorias podem ser consideradas. A utilização de dados reais de uma farmácia tornaria as análises mais precisas e aplicáveis ao contexto prático. Além disso, seria possível desenvolver painéis interativos, com gráficos dinâmicos e filtros, que permitiriam uma análise mais flexível e exploratória. Outra possibilidade seria aplicar modelos preditivos para estimar, por exemplo, quais produtos tendem a ter maior volume de vendas em determinados períodos. Também seria interessante segmentar os clientes por perfis de compra, visando compreender melhor seus comportamentos, além de analisar como promoções e métodos de pagamento impactam diretamente nas vendas. Todas essas melhorias tornariam a análise mais robusta, aplicável e próxima da realidade do mercado.

# Apêndices

Links referentes aos Colabs que contém as consultas.

Séries Temporais: https://colab.research.google.com/drive/1P-cWYl6P4LQCm99FgjELxCBDV0le6wmT#scrollTo=264cc732

Visualizações com Gráficos Estatísticos: https://colab.research.google.com/drive/1sBQKlezsmcTmrAxu50SyjonC\_KIWKrH5#scrollTo=e25eedf4

Medidas de Tendência Central (Média, Mediana e Moda):

https://colab.research.google.com/drive/1BhSfgUrlivRXUmzSO8Y4Om9\_5Vz\_wv2h?authuser=0

Medidas de Dispersão:

https://colab.research.google.com/drive/1iIt12ZnqzB6K9pHQ-U5XkopvGeBYN9jr?usp=sharing

Correlação entre Variáveis:

https://colab.research.google.com/drive/1MCL3WxJDAv4gnrcYLuEcBSd3Elw9bbgM?usp=sharing

Tipos de Amostragem: https://colab.research.google.com/drive/1gwf651hzNN-eZ5TG\_wCgniV5pkr9GfeT#scrollTo=siWCOCWO7t4E

# Referências

KHAN ACADEMY. Estatística e probabilidade. Disponível em:<https://pt.khanacademy.org/math/statistics-probability>. Acesso em: 21 jun. 2025.

DATA HACKERS. Disponível em:<https://www.datahackers.com.br/>. Acesso em: 21 jun. 2025.

TOWARDS DATA SCIENCE. Disponível em:<https://towardsdatascience.com/>. Acesso em: 21 jun. 2025.

PANDAS. Pandas Documentation. Disponível em:<https://pandas.pydata.org/docs/>. Acesso em: 21 jun. 2025.

SEABORN. Seaborn Documentation. Disponível em:<https://seaborn.pydata.org/>. Acesso em: 21 jun. 2025.

MATPLOTLIB. Matplotlib Documentation. Disponível em:<https://matplotlib.org/stable/index.html>. Acesso em: 21 jun. 2025.

NUMPY. NumPy Documentation. Disponível em:<https://numpy.org/doc/>. Acesso em: 21 jun. 2025.

SOUZA, André. **Séries Temporais Básicas**. GitHub, 2024. Disponível em:<https://github.com/profAndreSouza/Material/blob/main/Ci%C3%AAncia%20de%20Dados/08%20S%C3%A9ries%20Temporais%20B%C3%A1sicas.md>. Acesso em: 21 jun. 2025.

GOOGLE COLAB. **Análise de Séries Temporais – Código**. Disponível em:<https://colab.research.google.com/drive/1xVg3hfX4u4r6_yNopmKWOT0i-VwCBS1T>. Acesso em: 21 jun. 2025.

GOOGLE COLAB. **Análise de Correlação – Código**. Disponível em:<https://colab.research.google.com/drive/1S0auh-KWKe6i8He1S6GLPUhOBpczwfZO>. Acesso em: 21 jun. 2025.

SOUZA, André. **Dispersão dos Dados**. GitHub, 2024. Disponível em:<https://github.com/profAndreSouza/Material/blob/main/Ci%C3%AAncia%20de%20Dados/04%20Dispers%C3%A3o%20dos%20Dados.md>. Acesso em: 21 jun. 2025.

SOUZA, André. **Correlação e Associação**. GitHub, 2024. Disponível em:<https://github.com/profAndreSouza/Material/blob/main/Ci%C3%AAncia%20de%20Dados/09%20Correla%C3%A7%C3%A3o%20e%20Associa%C3%A7%C3%A3o.md>. Acesso em: 21 jun. 2025.

SOUZA, André. **Amostragem e Distribuições**. GitHub, 2024. Disponível em:<https://github.com/profAndreSouza/Material/blob/main/Ci%C3%AAncia%20de%20Dados/17%20Amostragem%20e%20Distribui%C3%A7%C3%B5es.md>. Acesso em: 21 jun. 2025.