

# Trabalho de Dados

*Ciência de Dados*

João Victor Machado  
Vitor Henrique Fogaça  
Pedro Vieira Cavalcium

Sorocaba  
2025

## Sumário

<b>Índice de Ilustração.....</b>	<b>5</b>
<b>Introdução.....</b>	<b>5</b>
<b>Objetivo ....Erro! Indicador não definido.</b>	
<b>Metodologia ..... Erro! Indicador não definido.</b>	
<b>Resultados Esperados.....</b>	<b>6</b>
<b>Anexos.....</b>	<b>6</b>
<b><u>Anexos</u>.....</b>	<b>6</b>

## Índice de Ilustração

Figura 1 - Análise de Dados .....	8
Figura 2 - Medidas de Tendência Central .....	8
Figura 3 - Distribuição das Notas .....	9
Figura 4 - Notas POr Gênero .....	9
Figura 5 - Normalidade .....	10

## Introdução

O avanço da tecnologia da informação e a crescente digitalização de processos educacionais têm impulsionado a geração massiva de dados no ambiente universitário. Diante desse cenário, torna-se fundamental o uso integrado de **Ciência de Dados** para transformar esses registros em conhecimento estratégico, capaz de subsidiar decisões acadêmicas e administrativas com maior precisão e eficiência.

Este projeto tem como foco o desenvolvimento de uma solução que una a robustez dos **Ciência de dados relacionais**, responsáveis pelo armazenamento estruturado e seguro das informações institucionais. A partir de dados simulados ou reais da universidade, como matrículas, desempenho dos alunos, taxas de evasão, frequência e histórico acadêmico, o projeto busca gerar relatórios e visualizações que facilitem a compreensão do comportamento estudantil e o aprimoramento da gestão acadêmica.

# Objetivo

## 2.1 Objetivo Geral

Desenvolver uma solução que integre técnicas de **Ciência de Dados** com foco na **normalização, estruturação, análise e visualização de dados universitários**, a fim de facilitar a identificação de padrões, desvios e informações relevantes para a gestão acadêmica.

## 2.2 Objetivos Específicos

- Aplicar as regras de **normalização** de dados para estruturar corretamente os dados acadêmicos da universidade, garantindo integridade, eficiência e escalabilidade do banco;
- Realizar a **limpeza e o tratamento dos dados** para garantir a qualidade e a consistência das análises;
- Utilizar técnicas de **Ciência de Dados** para explorar, analisar e interpretar os dados acadêmicos por meio de ferramentas como Python e bibliotecas (Pandas, Matplotlib, Seaborn) ou Power BI;
- Criar **gráficos e visualizações interativas** que destaquem informações relevantes, como desempenho dos alunos, evasão, frequência e correlações entre variáveis;
- Apontar **tendências, desvios e anomalias** nos dados que possam auxiliar na tomada de decisão por parte da administração universitária;
- Documentar todas as etapas do projeto, desde a modelagem do banco de dados até a análise dos resultados obtidos.

# Metodologia

A metodologia adotada neste projeto foi baseada na aplicação prática de conceitos de **Ciência de Dados**, utilizando ferramentas específicas para cada etapa do desenvolvimento. O projeto foi conduzido em fases sequenciais que envolveram desde a modelagem conceitual dos dados até a análise exploratória e visualização de informações relevantes para o contexto universitário.

## 3.3 Análise e Visualização de Dados

A etapa seguinte envolveu o uso de **Ciência de Dados** para analisar os dados armazenados. Utilizou-se o ambiente **Google Colab**, que permite a execução de scripts em Python de forma colaborativa e baseada na nuvem. Foram empregadas bibliotecas como **Pandas**, **Matplotlib** e **Seaborn** para:

- Limpeza e preparação dos dados extraídos do banco;
- Geração de gráficos e dashboards;
- Identificação de padrões, desvios e possíveis correlações entre variáveis acadêmicas (como desempenho, frequência e evasão).

## 3.4 Organização e Controle de Versões

Para garantir organização e controle das tarefas durante o desenvolvimento, foram utilizadas as plataformas **Git** e **GitHub**. A equipe manteve o código-fonte versionado, documentou as etapas do projeto e distribuiu as responsabilidades por meio de

issues, branches e commits, assegurando uma gestão colaborativa e eficiente das entregas.

## Resultados Esperados

O desenvolvimento deste projeto visa alcançar uma solução funcional e bem estruturada que una os conhecimentos de **Ciência de Dados** para gerar valor real a partir dos dados acadêmicos de uma universidade. A expectativa é que a plataforma criada contribua significativamente para a **tomada de decisões estratégicas e operacionais** por parte da instituição de ensino, com base em dados consistentes, organizados e visuais.

Os principais resultados esperados incluem:

- Uma **base de dados normalizada** e bem estruturada, com tabelas organizadas de forma relacional, representando fielmente os principais processos acadêmicos (como matrículas, disciplinas, notas, frequência, evasão, etc.);
- A **automação da coleta, limpeza e organização dos dados**, garantindo integridade e facilidade de análise;
- A geração de **visualizações claras e interativas**, como gráficos e dashboards, que permitam a identificação de:
  - Desvios de desempenho acadêmico;
  - Perfis de alunos com maior ou menor rendimento;
  - Possíveis correlações entre frequência e notas;
  - Padrões de evasão por curso ou período;

# Anexos

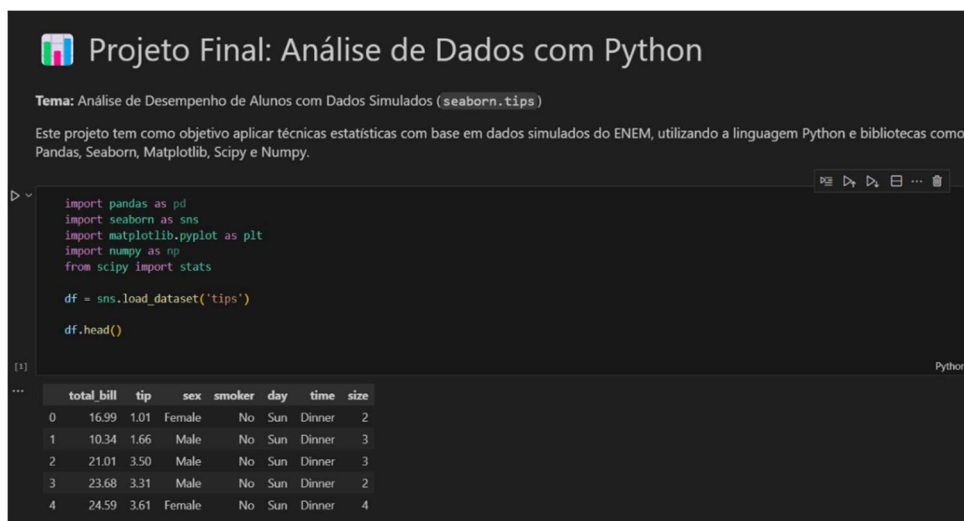


FIGURA 1 - ANÁLISE DE DADOS

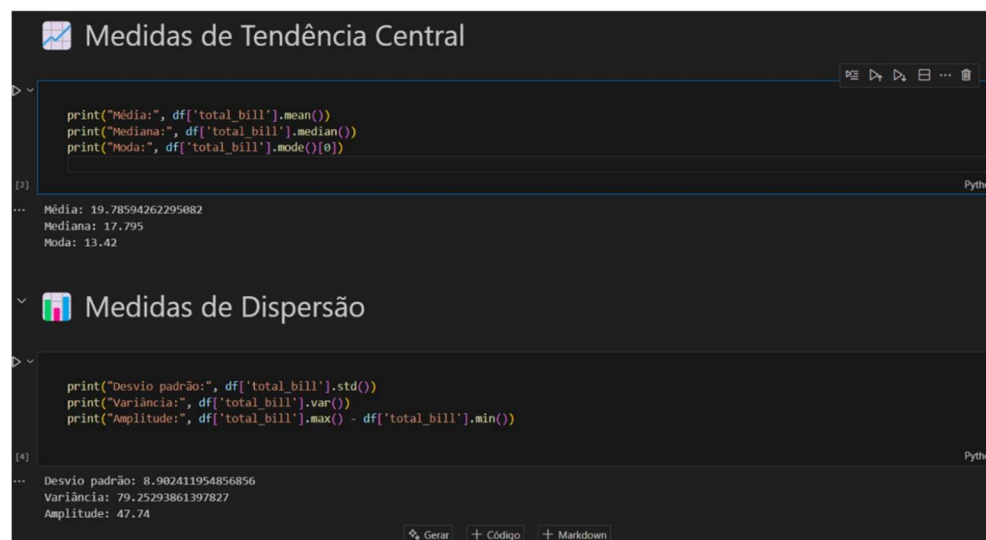


FIGURA 2 - MEDIDAS DE TENDÊNCIA CENTRAL



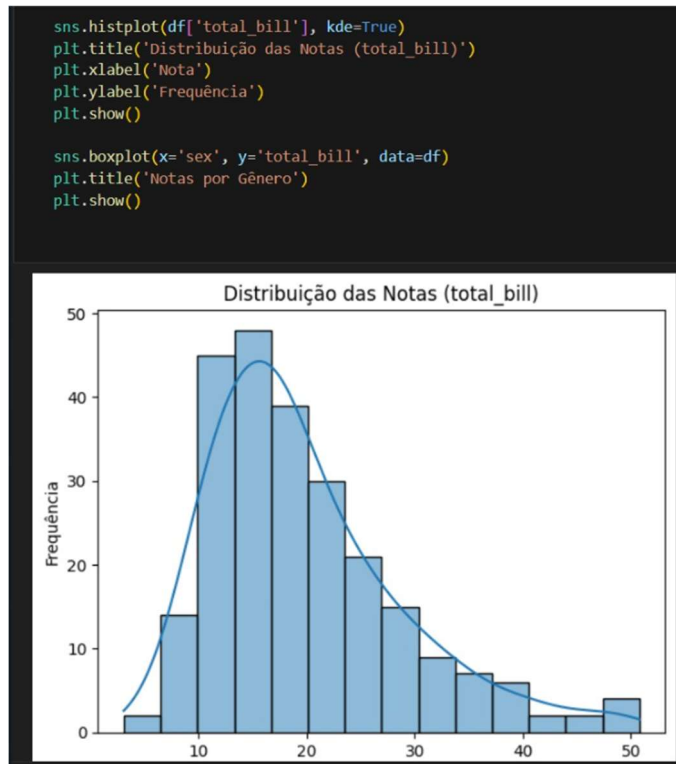


FIGURA 3 - DISTRIBUIÇÃO DAS NOTAS

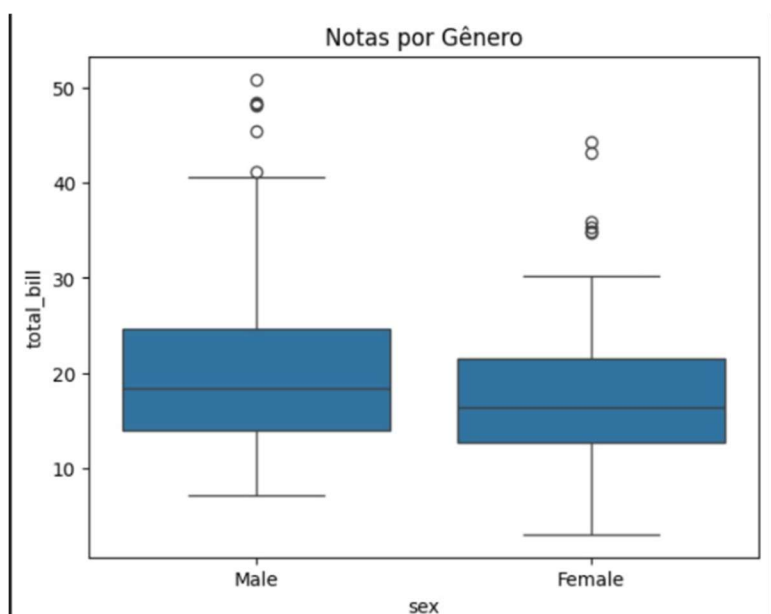


FIGURA 4 - NOTAS POR GÊNERO

## Teste de Normalidade (Shapiro-Wilk)

```
stat, p = stats.shapiro(df['total_bill'])
print("Estatística:", stat)
print("p-valor:", p)
if p > 0.05:
    print("Distribuição normal")
else:
    print("Distribuição não normal")
```

[6]

```
... Estatística: 0.9197187941346584
p-valor: 3.324539186809091e-10
Distribuição não normal
```

FIGURA 5 - NORMALIDADE

## Referências

11. IBM. *What is Data Science?* IBM, 2023. Disponível em: <https://www.ibm.com/br-pt/topics/data-science>. Acesso em: 6 jun. 2025.
12. W3Schools. *SQL Tutorial*. Disponível em: <https://www.w3schools.com/sql/>. Acesso em: 6 jun. 2025.
13. PostgreSQL Global Development Group. *PostgreSQL Documentation*. Disponível em: <https://www.postgresql.org/docs/>. Acesso em: 6 jun. 2025.