ESCOLA TÉCNICA SENAI GASPAR RICARDO JUNIOR

CURSO TÉCNOLOGO EM ANALISE DE DESENVOLVIMENTO E SISTEMAS

Felipe Cutiur Dos Santos

José Eduardo

Joacir Peçanha

Kelvim Lucas De Paula

**PROJETO FINAL:** Análise do Desempenho Escolar no Enem

Professor André Souza

Ciência de Dados

SOROCABA

2025

Sumário

[1. INTRODUÇÃO 5](#_Toc201349702)

[1.1. TEMA 5](#_Toc201349703)

[1.2. OBJETIVO DO PROJETO 5](#_Toc201349704)

[2. OBJETIVOS 5](#_Toc201349705)

[2.1. OBJETIVO GERAL DA PESQUISA 5](#_Toc201349706)

[2.2. OBJETIVOS ESPECÍFICOS DA PESQUISA 5](#_Toc201349707)

[3. METODOLOGIA APLICADA 5](#_Toc201349708)

[4. CIÊNCIA DE DADOS 6](#_Toc201349709)

[5. CONCLUSÃO 6](#_Toc201349710)

[6. REFÊRENCIAS 7](#_Toc201349711)

# INTRODUÇÃO

## TEMA

O exame nacional do ensino médio é a principal ferramenta de avaliação educacional no Brasil. Ele serve como parâmetro para que os alunos do ensino médio ingressem no ensino superior

## OBJETIVO DO PROJETO

Desafiar o conhecimento absorvido pelos membros do grupo durante o semestre referentes as Ciência de Dados, por meio de uma análise do desempenho dos estudantes no ENEM por meio de técnicas de análise estatística e visualização gráfica.

# OBJETIVOS

## OBJETIVO GERAL DA PESQUISA

O Objetivo geral da pesquisa se resume em analisar o desempenho geral dos alunos que prestaram o Exame Nacional do Ensino Médio (ENEM) ao longo dos anos, procurando detectar padrões e tendências de acordo com a região, tipo de escola e desempenho geral do aluno por disciplina.

## OBJETIVOS ESPECÍFICOS DA PESQUISA

Aplicar os conceitos de estatística aplicada à análise de dados aprendidos durante a disciplina em uma demanda real, identificando padrões e os representando por meio de gráficos que permitam a retirada de insights e fácil interpretação.

# METODOLOGIA APLICADA

3.1 BASE DE DADOS UTILIZADA

Utilizamos uma base de Dados própria que foi gerada com auxílio das bibliotecas Python Pandas e Numpy, contendo dados como notas simuladas, sexo, idade, escola e UF.

3.2 TECNOLOGIAS E FERRAMENTAS UTILIZADAS

Foi utilizada a linguagem Python em sua versão 3.11, se utilizando das bibliotecas voltadas para ciência de dados, como numPy, pandas, matplotlib, Seaborn e SciPy. A plataforma escolhida para a criação dos notebooks foi a Google Collab, por sua fácil acessibilidade, tempo de resposta e escalabilidade.

# CIÊNCIA DE DADOS

Os gráficos gerados utilizando as bibliotecas Matplotlib e Seaborn, estão disponíveis no seguinte link:

<https://github.com/ProjetoFinalAndre/Projeto-Final-CienciaDeDados/tree/pythoncode/graficos>

Tela de celular com texto preto sobre fundo branco

O conteúdo gerado por IA pode estar incorreto.

Figura 1 – Tabela Pandas com os dados.

Interface gráfica do usuário, Teams

O conteúdo gerado por IA pode estar incorreto.

Figura 2 – Aplicação de Amplitude, Desvio Padrão e Variância.

Interface gráfica do usuário, Tabela

O conteúdo gerado por IA pode estar incorreto.

Figura 3 – Média, Mediana e Moda a partir das notas.

Gráfico, Histograma

O conteúdo gerado por IA pode estar incorreto.

Figura 4 – Primeiro teste de normalidade.

Gráfico, Histograma

O conteúdo gerado por IA pode estar incorreto.

Figura 5 – Segundo teste de normalidade.

Gráfico, Gráfico de caixa estreita

O conteúdo gerado por IA pode estar incorreto.

Figura 6 – Gráfico Boxplot gerado analisando notas de acordo com o sexo dos candidatos.

Gráfico, Gráfico de caixa estreita

O conteúdo gerado por IA pode estar incorreto.

Figura 7 – Segundo gráfico Boxplot gerado analisando notas de acordo com o sexo dos candidatos.

Gráfico, Gráfico de dispersão

O conteúdo gerado por IA pode estar incorreto.

Figura 8 – Gráfico de dispersão entre as notas.

Interface gráfica do usuário, Aplicativo, Gráfico de funil

O conteúdo gerado por IA pode estar incorreto.

Figura 9 – Correlação de Pearson.

Link para o github com o Script Python completo:

<https://github.com/ProjetoFinalAndre/Projeto-Final-CienciaDeDados/blob/pythoncode/python/Codigos_ciencia_de_dados.ipynbCONCLUSÃO>

1. CONCLUSÃO

A pesquisa permitiu identificarmos padrões evidentes no desempenho dos estudantes que realizaram o Exame Nacional do Ensino Médio (ENEM). Destacando a superioridade de médias nas notas de redação, uma maior variância nas Ciências Exatas e suas tecnologias e um desempenho superior em candidatos do sexo feminino na Área de Linguagens e suas tecnologias, o uso do Python como linguagem principal para análise de dados e análise estatística novamente se mostrou extremamente eficiente para retirar insights relevantes de grandes conjuntos de dados.

# REFÊRENCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 14724:2011 – Informação e documentação — Trabalhos acadêmicos — Apresentação. Rio de Janeiro, 2011.

SCIPY. *Scientific Library for Python*. Disponível em: <https://scipy.org/>

PANDAS. *Python Data Analysis Library*. Disponível em: <https://pandas.pydata.org/>

SEABORN. *Statistical Data Visualization*. Disponível em: <https://seaborn.pydata.org/>

GOOGLE COLAB. *Google Collaboratory*. Disponível em: <https://colab.research.google.com/>