

Relatório Final: Análise de Dados Educacionais e ENEM 2022

Resumo Executivo

Este relatório apresenta uma análise abrangente do desempenho dos estudantes brasileiros no ENEM 2022, correlacionando-o com fatores como infraestrutura escolar e indicadores socioeconômicos dos municípios. Os dados foram extraídos de três fontes principais: Microdados do ENEM 2022, Censo Escolar 2022 e Indicadores Socioeconômicos por Município. A análise revela padrões significativos que podem orientar políticas educacionais e investimentos para a melhoria da qualidade da educação no Brasil.

1. Introdução

O Exame Nacional do Ensino Médio (ENEM) é a principal porta de entrada para o ensino superior no Brasil e um importante indicador da qualidade da educação básica. Neste estudo, investigamos como diferentes fatores influenciam o desempenho dos estudantes, buscando identificar pontos de intervenção para reduzir desigualdades educacionais.

1.1 Objetivos da Análise

- Identificar padrões de desempenho no ENEM 2022 por região, tipo de escola e perfil socioeconômico
- Analisar a relação entre infraestrutura escolar e resultados acadêmicos
- Investigar como o IDH dos municípios se relaciona com o desempenho dos estudantes
- Explorar diferenças de desempenho por faixa etária e gênero

1.2 Bases de Dados Utilizadas

- Microdados do ENEM 2022** (INEP): Resultados individuais dos estudantes, incluindo notas por área de conhecimento e dados socioeconômicos
- Censo Escolar 2022** (INEP): Informações sobre infraestrutura, docentes e matrículas das escolas brasileiras
- Indicadores Socioeconômicos por Município** (IBGE): Dados de IDH, PIB per capita e outros indicadores municipais

2. Metodologia

2.1 Tratamento dos Dados

O processo de tratamento dos dados incluiu:

- Limpeza e padronização de formatos
- Tratamento de valores ausentes
- Criação de colunas derivadas para análises específicas:
 - Média geral das notas por participante
 - Categorização em faixas etárias
 - Índice de infraestrutura escolar
 - Categorização do IDH municipal

2.2 Técnicas de Análise

- Análise descritiva das distribuições de notas
- Comparações entre grupos (escola pública vs. privada, faixas etárias)
- Análise de correlação entre variáveis
- Visualizações geoespaciais

3. Principais Descobertas

3.1 Panorama Geral do Desempenho

A distribuição das notas apresenta padrões distintos entre as diferentes áreas de conhecimento:

- **Linguagens e Códigos:** Distribuição mais homogênea, com média nacional aproximada de 520 pontos
- **Matemática:** Maior variabilidade, com polarização entre baixas e altas notas
- **Ciências da Natureza:** Tendência central mais baixa, indicando maior dificuldade
- **Ciências Humanas:** Distribuição ligeiramente assimétrica à direita
- **Redação:** Distribuição multimodal, sugerindo diferentes níveis de habilidade na escrita

3.2 Diferenças por Tipo de Escola

A análise revelou disparidades significativas entre escolas públicas e privadas:

- Estudantes de escolas privadas obtiveram, em média, notas 20% superiores aos de escolas públicas
- A maior diferença foi observada em Matemática, seguida por Ciências da Natureza
- A menor diferença foi observada em Linguagens e Códigos
- Escolas privadas apresentam, em média, índice de infraestrutura 40% superior às públicas

3.3 Relação entre Infraestrutura e Desempenho

Foi identificada uma correlação positiva moderada ($r = 0.42$) entre o nível de infraestrutura escolar e o desempenho dos estudantes:

- Escolas com infraestrutura classificada como "Avançada" apresentaram média de notas 15% superior às de infraestrutura "Básica"
- A presença de laboratórios de ciências e informática mostrou-se particularmente relevante para o desempenho em Ciências da Natureza e Matemática
- O impacto da infraestrutura é mais pronunciado em escolas públicas do que em privadas

3.4 Influência de Fatores Socioeconômicos

Os indicadores socioeconômicos dos municípios mostraram forte associação com o desempenho:

- Municípios com IDH classificado como "Muito alto" apresentaram média de notas 25% superior aos de IDH "Baixo"
- A correlação entre PIB per capita e desempenho é positiva, mas não linear, sugerindo que outros fatores interferem nessa relação
- A desigualdade de renda dentro dos municípios mostrou correlação negativa com o desempenho médio

3.5 Análise por Faixa Etária e Gênero

- O desempenho tende a ser melhor na faixa etária de 18 a 20 anos, declinando nas faixas etárias mais elevadas
- Estudantes do sexo feminino apresentaram melhor desempenho em Linguagens, Códigos e Redação
- Estudantes do sexo masculino obtiveram notas ligeiramente superiores em Matemática e Ciências da Natureza
- A diferença de desempenho por gênero é menos pronunciada em escolas com melhor infraestrutura

4. Conclusões e Recomendações

4.1 Principais Insights

1. **Infraestrutura importa:** Há evidências claras de que o investimento em infraestrutura escolar está associado a melhores resultados acadêmicos, especialmente em escolas públicas.
2. **Desigualdade regional persiste:** As disparidades de desempenho entre regiões refletem desigualdades socioeconômicas mais amplas, exigindo políticas compensatórias.
3. **Tipo de escola é determinante:** A diferença de desempenho entre escolas públicas e privadas permanece significativa, mesmo controlando por fatores socioeconômicos.
4. **Matemática como gargalo:** O desempenho em Matemática apresenta maior variabilidade e as maiores disparidades entre grupos, sugerindo necessidade de atenção especial a esta área.
5. **IDH como preditor:** O IDH municipal mostrou-se um bom preditor do desempenho médio, indicando a importância de políticas integradas de desenvolvimento humano.

4.2 Recomendações

Com base nas análises realizadas, recomendamos:

1. **Investimento direcionado em infraestrutura:** Priorizar escolas em municípios de baixo IDH, com foco em laboratórios de ciências e informática.
2. **Programas específicos para Matemática:** Desenvolver iniciativas focadas na melhoria do ensino de Matemática, especialmente na rede pública.
3. **Políticas intersetoriais:** Integrar políticas educacionais com outras políticas de desenvolvimento social para amplificar resultados.
4. **Formação docente:** Investir na qualificação de professores, com ênfase nas áreas com maior disparidade de resultados.
5. **Monitoramento contínuo:** Implementar sistemas de monitoramento que permitam acompanhar a evolução do desempenho e ajustar intervenções.

5. Limitações do Estudo e Pesquisas Futuras

5.1 Limitações

- Análise baseada em dados de um único ano, não captando tendências temporais
- Impossibilidade de estabelecer relações causais definitivas
- Dados autodeclarados podem conter imprecisões
- Amostra pode não representar adequadamente estudantes que não realizaram o ENEM

5.2 Sugestões para Pesquisas Futuras

- Análise longitudinal, acompanhando a evolução do desempenho ao longo dos anos
- Estudos de caso em escolas que obtiveram resultados acima do esperado para seu contexto socioeconômico
- Investigação mais detalhada do impacto de programas específicos de política educacional
- Análise da trajetória dos estudantes após o ENEM, correlacionando com seu desempenho

Anexos

1. <https://lookerstudio.google.com/s/r1ldwMqJeWw> (<https://lookerstudio.google.com/s/r1ldwMqJeWw>).
2. Documentação técnica do tratamento dos dados
3. Código-fonte das análises realizadas
4. Visualizações complementares

Este relatório foi elaborado como parte do Projeto Final do Curso de Analista de Dados.

Data de finalização: 03/04/2025

