

**Tecnologia pela Equidade Educacional no Brasil: Desenvolvimento de Ferramenta para Democratizar a Análise de Dados**

*Technology for Educational Equity in Brazil: Development of a Tool to Democratize Data Analysis*

*Tecnología Para La Equidad Educativa En Brasil: Desarrollo De Una Herramienta Para Democratizar El Análisis De Datos*

Allan Coelho Monteiro<sup>1</sup>

Lucas Machado da Silva Santana<sup>2</sup>

Ronilson Rodrigues Pinho<sup>3</sup>

Sildenir Alves Ribeiro<sup>4</sup>

**Resumo.** O presente estudo analisou, sob uma ótica quantitativa, os microdados do Exame Nacional do Ensino Médio (Enem) ao longo de dez anos (2014–2023), com o objetivo de investigar a equidade educacional no Brasil. Para embasar o presente estudo, foi realizado um levantamento bibliográfico da literatura, selecionando pesquisas que analisassem desempenho e desigualdades educacionais no Brasil, visando identificar e superar os desafios técnicos na manipulação e análise desse grande volume de dados, identificado na literatura. Além disso, foi desenvolvida uma aplicação web interativa que facilita a visualização e interpretação desses dados, e ainda permite que comparações entre grupos socioeconomicamente e geográficos distintos, sejam realizadas. A ferramenta foi construída com Python, HTML/CSS, e o framework Dash, além das bibliotecas Plotly e Pandas. Os resultados confirmam a influência determinante que variáveis socioeconômicas como renda familiar, cor/raça, tipo de escola e região exercem sobre o desempenho dos estudantes, revelando a existência de padrões persistentes de desigualdade interseccional. A pesquisa foi guiada por duas hipóteses: H1 — A integração dos dados do Enem, entre 2014 a 2023, e de todos os estados brasileiros revelam padrões de desigualdade interseccional e tendências de longo prazo? H2 — A aplicação desenvolvida possibilita análises eficientes e democratiza o acesso a informações para pesquisas educacionais? Conclui-se que a aplicação não apenas valida as hipóteses de pesquisa, mas também se mostra uma contribuição metodológica significativa ao superar as barreiras na manipulação de grandes volumes de dados educacionais, democratizando o acesso e a análise a informações essenciais para pesquisas e políticas públicas que se baseiam em evidências.

**Palavras-chave:** Equidade Educacional. Desigualdade educacional. Enem. Aplicação Web. Dashboard.

**Abstract.** The present study analyzed, from a quantitative perspective, the microdata of the National High School Examination (Enem) over a ten-year period (2014–2023), with the aim of investigating educational equity in Brazil. To support this study, a bibliographic review of the literature was carried out, selecting research that analyzed educational performance and inequalities in Brazil. In order to overcome the technical challenge of handling and analyzing this large volume of data—identified in the literature—we developed an interactive web application that facilitates the visualization and interpretation of these data, while also enabling comparisons between socioeconomically and

---

<sup>1</sup> Bacharel em Sistemas de Informações. CEFET/RJ. <https://orcid.org/0009-0001-2207-0339>. E-mail: allan.dev82@gmail.com

<sup>2</sup> Bacharel em Sistemas de Informações. CEFET/RJ. <https://orcid.org/0009-0006-1946-5280>. E-mail: lucasmssantana@gmail.com

<sup>3</sup> Mestre em Modelagem Matemática e Computacional. CEFET/RJ. <https://orcid.org/0009-0008-0852-2803>. E-mail: ronilson.pinho@cefet-rj.br

<sup>4</sup> Doutor em Informática. CEFET/RJ. <https://orcid.org/0000-0003-4808-1009>. E-mail: sildenir.ribeiro@cefet-rj.br

geographically distinct groups. The tool was built with Python, HTML/CSS, and the Dash framework, along with the Plotly and Pandas libraries. The results confirm the decisive influence of socioeconomic variables such as family income, race/color, school type, and region on students' performance, revealing the existence of persistent patterns of intersectional inequality. The research was guided by two hypotheses: H1 — Does the integration of Enem data, from 2014 to 2023, across all Brazilian states reveal patterns of intersectional inequality and long-term trends? H2 — Does the developed application enable efficient analyses and democratize access to information for educational research? It is concluded that the application not only validates the research hypotheses but also represents a significant methodological contribution by overcoming barriers in handling large volumes of educational data, democratizing access to and analysis of essential information for evidence-based research and public policies.

**Keywords:** Educational Equity. Educational Inequality. Enem. Web Application. Dashboard.

**Resumen.** El presente estudio analizó, bajo una óptica cuantitativa, los microdatos del Examen Nacional de la Enseñanza Media (Enem) a lo largo de diez años (2014–2023), con el objetivo de investigar la equidad educativa en Brasil. Para fundamentar el presente estudio, se realizó un levantamiento bibliográfico de la literatura, seleccionando investigaciones que analizaran el desempeño y las desigualdades educativas en Brasil. Con el fin de superar el desafío técnico de la manipulación y el análisis de este gran volumen de datos, identificado en la literatura, desarrollamos una aplicación web interactiva que facilita la visualización e interpretación de dichos datos, y que además permite realizar comparaciones entre grupos socioeconómica y geográficamente distintos. La herramienta fue construida con Python, HTML/CSS y el framework Dash, además de las bibliotecas Plotly y pandas. Los resultados confirman la influencia determinante que variables socioeconómicas como ingreso familiar, color/raza, tipo de escuela y región ejercen sobre el desempeño de los estudiantes, revelando la existencia de patrones persistentes de desigualdad interseccional. La investigación fue guiada por dos hipótesis: H1 — ¿La integración de los datos del Enem, entre 2014 y 2023, y de todos los estados brasileños revela patrones de desigualdad interseccional y tendencias a largo plazo? H2 — ¿La aplicación desarrollada posibilita análisis eficientes y democratiza el acceso a informaciones para investigaciones educativas? Se concluye que la aplicación no solo valida las hipótesis de investigación, sino que también se muestra como una contribución metodológica significativa al superar las barreras en la manipulación de grandes volúmenes de datos educativos, democratizando el acceso y el análisis de informaciones esenciales para investigaciones y políticas públicas basadas en evidencias.

**Palabras clave:** Equidad Educativa. Desigualdad educativa. Enem. Aplicación Web. Dashboard.

## Introdução

Na última edição do Programa Internacional de Avaliação de Estudantes (*PISA – Program for International Student Assessment*) no Brasil, em 2022, o país ficou abaixo da média da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) em todos os testes, ficando atrás de Costa Rica, México, Uruguai, Peru, Argentina e Colômbia. De oitenta e um países participantes, o Brasil em matemática ficou na 65ª posição (Inep, 2022). Segundo relatório do *PISA*, divulgado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), 73% dos estudantes brasileiros não alcançaram o nível básico (nível 2) em matemática, considerado pela OCDE o mínimo necessário para que os jovens possam exercer plenamente sua cidadania. Alunos de 15 anos não sabem resolver problemas matemáticos simples, e estão tendo o aprendizado esperado para estudantes de 11 e 12 anos

(Inep, 2022; Futura, 2024). O *PISA* é um programa desenvolvido pela OCDE, que tem sua aplicação no Brasil desde 2000. É um dos programas de avaliação internacional mais desenvolvido e aplicado no mundo, reconhecido em nível internacional como produtor da maior quantidade de conhecimento sobre o sistema educativo de muitos países no mundo (Villani; Oliveira, 2018).

Enquanto os resultados do *PISA* evidenciam não apenas as deficiências no ensino brasileiro, mas refletem um cenário mais amplo de desigualdades estruturais no sistema educacional, ressaltando a importância de instrumentos nacionais capazes de mapear disparidades socioeconômicas e regionais com maior precisão, o Exame Nacional do Ensino Médio (Enem), nesse sentido, se destaca como uma ferramenta estratégica para avaliação educacional. Diferente do *PISA*, que utiliza uma pequena amostra de estudantes, o Enem abrange milhões de participantes anualmente, coletando dados detalhados como renda, raça, tipo de escola, escolaridade dos pais, e localização geográfica (Inep, 2022; Brasil, 2025). Essa superioridade na produção de informação se torna mais evidente com a seguinte comparação: o *PISA* revela que 73% dos estudantes não atingiram o nível básico em matemática, mas o Enem permite identificar que alunos de escolas públicas no Norte têm notas médias 23,31% inferiores às dos alunos de escolas privadas (Rocha *et al.*, 2022), dado essencial para entender as raízes das desigualdades.

Devido a essa riqueza de dados gerados pelo Enem, o exame passou a exercer papel de destaque em pesquisas e debates que abordam o tema equidade educacional nos diferentes níveis educacionais, pois somos um país marcado por desigualdades na educação, que ora se apresenta na falta de acesso à educação, ora pela exclusão dentro do próprio sistema, ou ainda, na falta de padrões educacionais de qualidade (Sampaio; de Oliveira, 2015). Desigualdades que contrariam nossas leis, pois o acesso à educação é um direito de todos, garantido pela Constituição Federal do Brasil de 1988 (CF/88), previsto nos artigos 205 e 206, que garantem a todos o acesso à educação de qualidade (Brasil, 1988). Além disso, os dados gerados pelo Enem também contribuem com as metas do Plano Nacional de Educação (PNE), que busca garantir a todos o acesso à educação de qualidade e igualitária. Ele estabelece metas, objetivos e estratégias para a educação brasileira, e é renovado a cada dez anos (Brasil, 1996; Brasil, 2001). Atualmente estamos na vigência do PNE sancionado pela lei 13.005/2014 (Brasil, 2014).

Mas apesar do presente potencial dos microdados do Enem para a análise das desigualdades educacionais, a complexidade e o grande volume desses dados representam um desafio técnico significativo para pesquisadores e formuladores de políticas públicas. Com isso,

a necessidade de ferramentas que superem essa barreira metodológica e que democratizem o acesso a essas importantes informações é a principal lacuna de pesquisa que este estudo busca preencher. O presente trabalho desenvolveu uma ferramenta interativa para análise dos microdados do Enem (2014-2023) com o objetivo de facilitar a investigação das desigualdades, oferecendo uma contribuição metodológica para o campo da pesquisa educacional, e auxiliando, de maneira assertiva, na formulação de políticas públicas. Em um país como o Brasil, marcado por profundas desigualdades, essa pesquisa se torna relevante para garantir que o acesso à educação qualidade alcance a todos.

Já o termo “equidade educacional” tem se tornado relevante como tema de estudos sobre educação no Brasil. E na literatura, a noção de equidade educacional, considera o acesso, a permanência, aprendizagem, recursos, reconhecimento e diversidade (Barbosa et al., 2022). Mesmo reconhecendo que o acesso e a permanência são os pontos com mais relevância para o alcance da equidade (Dubet, 2008), nesse estudo equidade educacional será considerada a luz da justiça social, o que resgata o conceito original da palavra, oriundo da Grécia antiga. A palavra equidade, etimologicamente, advém do grego antigo *aequus*, que significa justo, sendo correntemente empregada para denotar igualdade e justiça (Portal Insights, 2024). Sendo assim, a equidade educacional pode ser entendida como uma aplicação de justiça corretiva, visando uma distribuição de forma equilibrada dos conhecimentos adquiridos entre os diferentes grupos de alunos (Crahay, 2000), tendo por objetivo diminuir as distâncias que estejam relacionadas com a origem social, condições econômicas e diferenças culturais (Soares; Delgado, 2016).

A busca por esse ideal à esteira da justiça social, fortalece a necessidade de se reconhecer as determinantes sociais de forma individualizada, para que a equidade educacional seja alcançada, tendo em vista que a meritocracia praticada na sociedade contemporânea não é perfeitamente adequada, pois o esforço dos indivíduos menos privilegiados nem sempre é compensado, tornando o acesso aos recursos uma corrida desigual (Ribeiro, 2014). Tal fato, corrobora para a importância da distinção entre a equidade educacional, da mera igualdade formal, o que pressupõe a distribuição diferenciada de recursos para garantir oportunidades reais a grupos desfavorecidos (Aldridge; Fraser, 2008). Segundo Piovesan (2005) igualdade formal, deve ser reduzida a fórmula “todos são iguais perante a lei”, que no seu tempo foi fundamental para a eliminação dos privilégios. Porém, igualdade material, é correspondente ao ideal de justiça social e distributiva (equidade), ou seja, uma igualdade orientada pelo critério socioeconômico, que corresponde ao ideal de justiça como reconhecimento de raça, idade, gênero, etnia e demais critérios. Assim, a justiça social como redistribuição e reconhecimento transita da igualdade formal para a igualdade material.

Nesse sentido, o presente estudo, visando contribuir de forma relevante com o tema, debruçou-se sobre duas hipóteses: H1 — A integração dos dados do Enem, entre 2014 a 2023, e de todos os estados brasileiros revelam padrões de desigualdade interseccional e tendências de longo prazo; H2 — A aplicação desenvolvida possibilita análises eficientes e democratiza o acesso a informações para pesquisas educacionais.

## Trabalhos Correlatos

Para embasar o presente estudo, foi realizado um levantamento bibliográfico da literatura, selecionando pesquisas que analisassem desempenho e desigualdades educacionais no Brasil, usando microdados do Enem. Assim como, trabalhos que discutiram políticas públicas com base em evidências quantitativas. A seleção dos trabalhos foi baseada em sua relevância para os objetivos deste estudo, priorizando pesquisas com abordagens metodológicas similares ou complementares. As buscas foram conduzidas nas plataformas Google Scholar e SciELO, e em bases de dados nacionais, utilizando *String* de busca elaboradas a partir de palavras-chave diretamente relacionadas ao tema.

Isso permitiu identificar lacunas existentes na literatura, e a necessidade de ferramentas tecnológicas que possibilitem uma análise integrada e acessível dos dados educacionais brasileiros. Os trabalhos selecionados servem como base para o desenvolvimento do presente estudo, que busca avançar nas discussões sobre a equidade educacional (Tabela 1).

Tabela 1 – Tabela Síntese

Autor(es)	Tema Central	Ano	Base de Dados	Contribuição
Santos	Desigualdades regionais no Enem	2019	Enem 2018	Background familiar impacta desempenho
Rocha et al.	Matemática no Enem (Nordeste)	2022	Enem 2017–2019	Diferenças por rede de ensino e renda
Ribeiro; De Oliveira	Desigualdade educacional e acesso ao ensino	2024	Dados oficiais e estatísticos	Disparidades regionais e raciais afetam escolarização e desempenho
Azevedo et al.	Desenvolvimento de dashboard para avaliação docente	2024	Relatório de Atividades Docentes (RAD) da UPE	Gestão universitária com visualizações interativas, facilita a tomada de decisões

Lucena; Dos Santos	Influências socioeconômicas no Enem	2020	Enem 2016	Reforça a literatura sobre o impacto socioeconômico no exame
Nakazone e Bortolotti	Análise histórica do Enem com Python e Pandas	2021	Enem 2015–2019	Identifica padrões regionais e socioeconômicos no desempenho.

Fonte: Autores (2025).

Os efeitos das desigualdades regionais nos resultados do Enem: uma análise a partir dos microdados de 2018

Santos (2019), investiga desigualdades regionais nos desempenhos do Enem de 2018, e verifica se fatores de background familiar como renda e escolaridade dos pais, impactam os resultados dos candidatos. A pesquisa utilizou o Método dos Mínimos Quadrados Ordinários (MQO) para estimar equações baseadas nos microdados do Enem 2018. Os resultados indicaram que a região de origem do indivíduo explica parte significativa do desempenho no Enem. E fatores como renda familiar e escolaridade da mãe mostraram-se determinantes, evidenciando desigualdades sociais e regionais. Por exemplo, regiões com maior renda per capita e maior escolaridade materna apresentaram melhores desempenhos médios.

Análise dos microdados de matemática do Enem de 2017–2019 do Nordeste

Rocha *et al.* (2022), publicado na revista *Research, Society and Development* em 2022, realizaram um estudo que investiga as desigualdades educacionais na região Nordeste do Brasil, analisando os microdados de desempenho dos estudantes em Matemática no Enem, durante o período de 2017 a 2019. A pesquisa foi motivada pela necessidade de compreender como fatores socioeconômicos e demográficos impactam o desempenho escolar, com o intuito de subsidiar políticas educacionais mais eficazes. A metodologia adotada combina pesquisa bibliográfica e documental, e análise dos microdados fornecidos pelo Inep. Ferramentas estatísticas foram empregadas para identificar padrões e correlações entre as variáveis analisadas. Os resultados revelam uma relação significativa entre o desempenho em Matemática e fatores socioeconômicos. Estudantes de escolas privadas e de famílias com maior renda apresentaram, em média, melhores resultados em comparação com aqueles provenientes de escolas públicas e contextos socioeconômicos mais vulneráveis.

Avanços e desafios: o caminho da Educação Brasileira rumo à equidade

A pesquisa elaborada por Ribeiro; De Oliveira (2024), publicada na Revista Observatório De La Economia Latino-americana, investiga os desafios contínuos de desigualdade no sistema educacional brasileiro, analisando os elementos que perpetuam diferenças regionais e socioeconômicas no acesso ao ensino de qualidade. Abordando de forma qualitativa e quantitativa, com revisão de documentos oficiais, dados estatísticos, e literatura especializada, questiona de que modo a má distribuição de investimentos, equipamentos e estrutura escolar influencia as chances de aprendizagem entre grupos sociais distintos no Brasil. Resultados como o de 2023, que demonstram o baixo acúmulo de escolarização, onde 42,4% da população com 25 anos ou mais não tinha ensino fundamental completo, sendo esse percentual ainda maior entre pessoas pretas ou pardas (54,2%), e pessoas residentes nas regiões Norte e Nordeste, 53,1% e 48,8% respectivamente. E o de desempenho na conclusão do ensino médio no Brasil em 2021, que teve uma taxa de aprovação de 77,3%, sendo inferior nas escolas públicas (74,4%) e ainda mais baixa nas regiões Norte (71,2%) e Nordeste (72,6%), enfatizam a persistência de disparidades significativas que afetam o acesso e a qualidade da educação em diferentes regiões e estratos sociais. Indicando que apesar dos avanços em termos de ampliação do acesso à educação básica e superior, as disparidades socioeconômicas continuam a impor barreiras significativas ao desenvolvimento educacional equitativo.

A relação entre desempenho no Exame Nacional do Ensino Médio e o perfil socioeconômico: um estudo com os microdados de 2016.

A pesquisa elaborada por Lucena e Santos (2020), publicada na Revista de Gestão e Secretariado (GeSec), investiga a influência de variáveis socioeconômicas no desempenho dos candidatos do Enem. Utilizando uma abordagem descritiva e quantitativa, com a análise de microdados do Enem de 2016 obtidos no Portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), o estudo questiona qual o impacto dessas características na nota geral dos estudantes. Os resultados demonstram que, em 2016, os candidatos com maior renda familiar e os que estudaram a maior parte do ensino médio em escolas particulares tendem a obter melhor desempenho na nota geral do Enem. Além disso, o estudo evidencia que candidatos que não trabalham e os de cor branca também apresentaram melhores resultados. A pesquisa aponta a persistência de disparidades significativas, corroborando estudos anteriores que abordaram análises semelhantes. Esses dados reforçam que as características socioeconômicas continuam a ser barreiras importantes para o desenvolvimento educacional equitativo no Brasil.

Concepção, construção e avaliação de um dashboard para apoio à tomada de decisão na Upe com base nos dados da avaliação de desempenho docente (RAD)

Azevedo et al. (2024), publicado na Jornada Científica e de Extensão da Universidade de Pernambuco (campus Caruaru), propõe o desenvolvimento e avaliação de um dashboard interativo com o objetivo de apoiar a tomada de decisão institucional a partir dos dados do Relatório de Atividades Docentes (RAD). O estudo foi motivado pela ausência de ferramentas visuais específicas para a análise desses dados. Sendo assim, o trabalho adota uma abordagem interativa fundamentada no *Design Activity Framework for Visualization Design*, integrado ao conceito de *Data-Informed Decision Making (DIDM)*. A ferramenta foi desenvolvida com o uso das seguintes ferramentas: Python, Plotly, Dash e Pandas. Foi avaliada com duas interações baseadas em entrevistas com gestores acadêmicos. O estudo contribuiu com soluções para a gestão da universidade, destacando a eficiência dos dashboard como instrumentos de apoio a soluções de análises. Entre os desafios identificados na pesquisa, destaca-se a necessidade de uma avaliação com mais usuários, e a implementação de novas funcionalidades, como agrupamento por unidades e métricas personalizáveis. Os resultados indicaram que o dashboard facilitou a compreensão dos dados e se mostrou útil para a tomada de decisões, permitindo análises temporais e comparações por unidades.

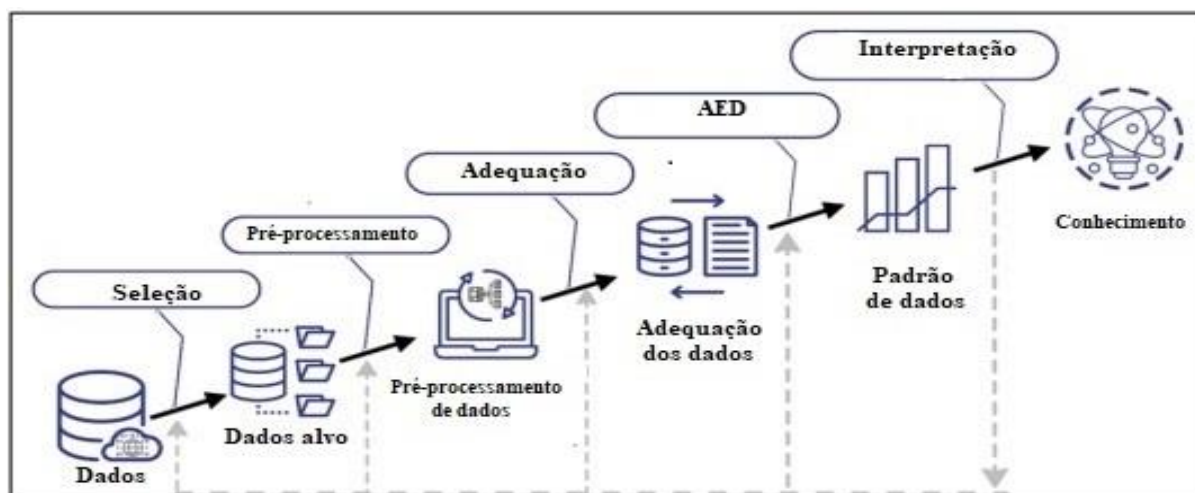
Análise de dados históricos do Enem 2015–2019 utilizando Python e Pandas

Nakazone e Bortolotti (2021) propõem uma análise dos microdados do Enem entre os anos de 2015 e 2019, utilizando ferramentas de ciência de dados como Python, Pandas, Matplotlib e Seaborn. A pesquisa está inserida no contexto do crescente volume de dados educacionais, destacando a importância da análise de grandes bases para a identificação de padrões e tendências que possam subsidiar melhorias no Sistema Educacional Brasileiro. A metodologia adotada envolve um estudo bibliográfico sobre Big Data e mineração de dados, seguido de um estudo de caso qualitativo com comparações entre os municípios de Mococa (SP), a região de Ribeirão Preto e a média nacional. Os resultados revelam correlações significativas entre desempenho no Enem e variáveis como renda familiar, região e tipo de escola, além de diferenças relevantes entre gêneros e impactos decorrentes de alterações no formato da prova. A contribuição do trabalho reside na geração de informações na área da educação que auxiliam na compreensão do desempenho estudantil em diferentes contextos. Entre as limitações apontadas, destaca-se a dificuldade de processamento de grandes volumes de dados, sugerindo a necessidade de abordagens mais especializadas em estudos futuros.

## Metodologia

Para esse estudo foi realizado inicialmente uma pesquisa bibliográfica nos documentos oficiais do exame na base de dados do Inep para que se pudesse entender a aplicação do exame e sua contribuição para a educação brasileira. Depois buscamos pesquisas científicas sobre os microdados do Enem nas bases de dados *Google Scholar e Scientific Electronic Library Online* (SciELO) para conhecermos a trajetória do exame, assim como, suas contribuições para o tema.

A pesquisa foi delineada como descritiva quanto aos objetivos, de fonte de dados como documental e de abordagem quantitativa (Prodanov; Freitas, 2013). A escolha da metodologia, incluindo as ferramentas e linguagens utilizadas, foi feita com o intuito de resolver os desafios da pesquisa, seguindo um panorama de processos que inclui seleção, dados alvo, pré-processamento, adequação e análise exploratória dos dados, resultando na interpretação e geração de conhecimento, conforme figura 1, que detalha essas etapas.



**Figura 1.** Panorama dos Processos –  
Fonte: Adaptado Dos Santos *et al.* (2024)

### Base de Dados e Processamento

Os dados utilizados neste estudo são públicos, obtidos diretamente do Inep, sendo assim, não requerem tratamento no que diz respeito a Lei nº 13.709/2018 (Lei Geral de Proteção de Dados Pessoais – LGPD). Esta base é considerada uma das mais relevantes para políticas públicas educacionais (Ferreira *et al.*, 2021), com dados consistentes, confiáveis, diversificados e numerosos (Rodrigues *et al.*, 2019). Foram considerados para o trabalho os microdados do Enem dos últimos dez anos disponíveis (2014 a 2023) na data de elaboração deste trabalho. Os arquivos originais somavam aproximadamente 40 GB, contendo

informações detalhadas sobre o desempenho, localização geográfica, e perfil socioeconômico, somando ao todo, 78 variáveis, de mais de 58 milhões de candidatos.

Em uma rigorosa etapa de pré-processamento, os dados foram tratados e adequados para a análise. Primeiro, foram removidos registros de candidatos que não compareceram às provas, ou que não preencheram o questionário socioeconômico relevante para o estudo. Essa etapa resultou na redução total de registros para 37 milhões, representando uma diminuição de 36,2%. Em seguida, as colunas não consideradas para o estudo foram eliminadas, passando de 78 para 13 variáveis, sendo a coluna “média geral” adicionada ao *dataframe*, pois não existia na base original. São elas: 'NU\_ANO', 'TP\_SEXO', 'TP\_COR\_RACA', 'TP\_ESCOLA', 'SG\_UF\_PROVA', 'NU\_NOTA\_CN', 'NU\_NOTA\_CH', 'NU\_NOTA\_LC', 'NU\_NOTA\_MT', 'NU\_NOTA\_REDACAO', 'MEDIA\_GERAL', 'Q005', 'Q006'.

Após a limpeza, os dados foram padronizados. A coluna 'Q006', que originalmente era tipo “*string*”, foi convertida para o tipo numérico ‘*int*’, garantido que as análises quantitativas fossem realizadas. Em seguida, os arquivos de cada ano, que originalmente estavam no formato CSV (*Comma-separated values*), foram unidos (concatenados) em um único arquivo, permitindo analisar tendência de longo prazo. E para otimizar a velocidade das consultas, o arquivo final foi convertido para a extensão *Apache Parquet*, que se mostrou mais eficiente que os formatos *CSV* e *JSON* (Microsoft, 2024). Todo esse processo permitiu uma redução do arquivo de aproximadamente 40 GB para 437 MB, totalizando ~99% do volume original.

### Ferramentas, Linguagens e Funcionamento da Aplicação

Para o desenvolvimento da análise e a construção da aplicação, foi utilizado a linguagem de programação python, com a distribuição Anaconda. Isso por ser uma linguagem muito utilizada em análises de dados, e ao mesmo tempo ser de propósito geral, ou seja, não é exclusiva para análises de dados como a linguagem R, o que permite utilizar a linguagem python para realizar análises, e ao mesmo tempo construir uma aplicação para apresentar os resultados dessas análises no mesmo ambiente de desenvolvido (Lopes *et al.*, 2019). As bibliotecas utilizadas nesse projeto, Pandas, Plotly e Dash, que são um conjunto de funções e módulos (Vicentiner *et al.*, 2020), foram empregadas para manipular os dados, criar gráficos e desenvolver interface web. Para a estruturação e o estilo da aplicação, as ferramentas HTML (HyperText Markup Language) e CSS (Cascading Style Sheets) foram utilizadas, permitindo customização, responsividade e personalizações avançadas. O Jupyter Notebook foi o IDE (IDE – *Integrated Development Environment*), por ser compatível com várias linguagens de

programação, e bastante utilizado na área de dados devido a sua flexibilidade, reprodutibilidade e integração com bibliotecas.

A aplicação foi projetada para ser interativa, permitindo que sejam feitas múltiplas seleções. O usuário pode selecionar o ano de referência e a área de análise. Em seguida, os parâmetros para os dois grupos (Grupo 1 e Grupo 2) podem ser definidos, incluindo estado, número de moradores, a renda da família, etnia (podendo ser mais de uma), sexo e tipo de escola. Após o processo de seleção, ao clicar em ‘Atualizar Gráficos’, os resultados são gerados, e exibidos através de gráficos boxplot para distribuição das notas de acordo com o ano selecionado, e gráficos de linha, contendo a média geral, para a evolução das médias ao longo dos dez anos.

### Critérios de Análises

Os critérios de análise foram definidos em três cenários principais para investigar diferentes aspectos de desigualdade.

Cenário 1 – Tipo de Escola (Pública vs. Privada) no estado do RJ: esta análise comparou o desempenho de 69,744 alunos de escolas públicas e privadas do Rio de Janeiro em 2023, e 1.044.118 de alunos para a evolução dos últimos dez anos. Em ambas o efeito do tipo de escola foi isolado, fixando as outras variáveis. O objetivo foi avaliar o desempenho das escolas através da distribuição das notas em 2023, e verificar sua evolução nos últimos 10 anos, sem incluir fatores qualitativos, como efeitos da pandemia e greves docentes.

Cenário 2 – Comparação Por Renda Familiar No Estado Do RJ: o foco desta seção foi avaliar o impacto da renda no desempenho acadêmico. Foram analisados 45.088 alunos, divididos em grupos com renda menor ou igual a dois salários-mínimos ( $\leq 2$  SM), e renda maior ou igual a dez salários-mínimos ( $\geq 10$  SM). E 660.187 alunos utilizados na análise da evolução dos últimos dez anos. Nos dois grupos o efeito renda familiar foi isolado através da fixação das outras variáveis, com o objetivo de avaliar o impacto da renda familiar no desempenho acadêmico.

Cenário 3 – Comparação Entre Estados Contrastantes: este cenário examinou disparidades estruturais ao cruzar fatores de etnia e região. Foram comparados 87.631 alunos brancos no Rio Grande do Sul, com alunos pretos e pardos da Bahia para a distribuição das notas em 2023, e 1.311.804 para a evolução das notas nos últimos dez anos. Essa escolha se justifica pelo fato de o Rio Grande do Sul ser o estado com maior proporção de brancos (78,4%), e a Bahia, o estado com a maior população de pretos e pardos (80,8%), conforme

dados do Censo do IBGE de 2022 (IBGE, 2022). O objetivo foi examinar como fatores étnicos e regionais influenciam o desempenho educacional.

## **Resultados e Discussão**

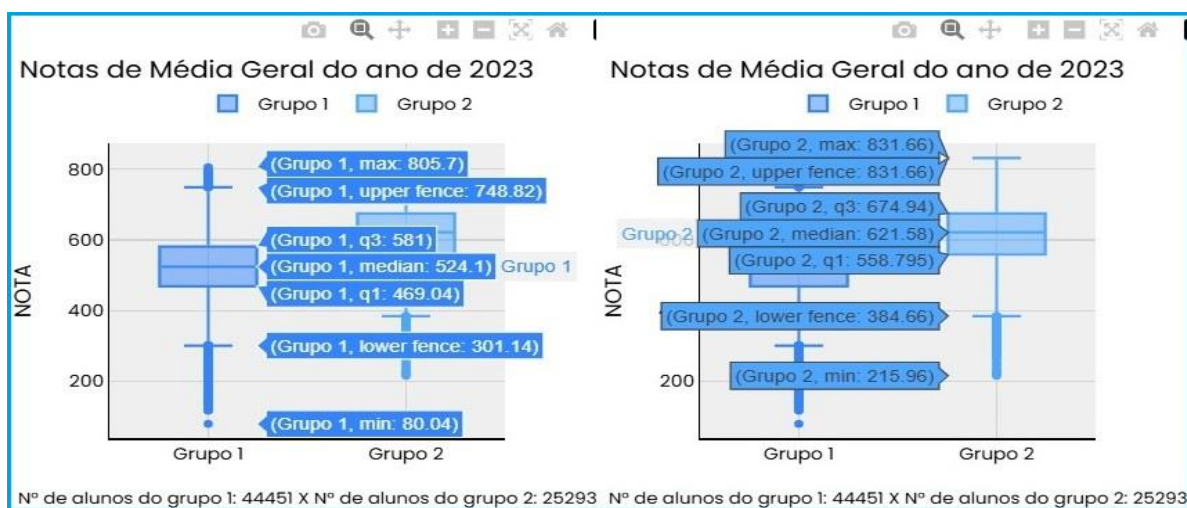
### Cenário 1 (Pública Vs. Privada)

O desempenho dos estudantes do ensino médio, conforme evidenciado pelos microdados do Enem, reflete as profundas desigualdades estruturais presentes no sistema educacional brasileiro, especialmente ao que tange à dependência administrativa das escolas, como observado em outros estudos (Rocha et al., 2022; Lucena; Dos Santos, 2020). Os resultados da presente pesquisa, ao analisar o Cenário 1, figura 2, corroboram com o que já foi afirmado na literatura, como no estudo de Rocha et al. (2022), que identificou que o desempenho no Enem de estudantes de escolas privadas, foi superior em relação aos de escolas públicas no Nordeste, entre 2017 e 2019. Segundo esse estudo as escolas públicas tiveram médias no intervalo de [300, 700), no entanto as escolas privadas estão concentradas na faixa de proficiência de [400, 800) e possuem a maior quantidade de participantes que alcançaram desempenho no intervalo [800, 1000). Nossas descobertas confirmam essa disparidade também no estado do Rio de Janeiro em 2023, mostrando que a mediana das notas em escolas privadas (Grupo 2), foi 97,48 pts (18,6%) superior à de escolas públicas (Grupo 1), mesmo controlando as variáveis demográficas.

Ainda constatamos que aproximadamente 40% dos candidatos de escola pública não conseguiram alcançar os 500 pontos, enquanto nas escolas privadas esse total não passou de  $\approx 15\%$ . O que corrobora com os resultados de Rocha et al. (2022) que em seu estudo demonstrou que 60,91% dos estudantes de escolas públicas tiveram desempenho menor ou igual a 500 pts, enquanto nas escolas privadas foi de 24,43%. Esse estudo ainda demonstrou que 75,57% dos alunos de escolas privadas obtiveram pontuação maior ou igual a 500 pts, contra 38,85% das redes públicas. Resultados próximos aos encontrados em nossa pesquisa, que identificou que 75% dos alunos de escolas privadas conseguiram notas igual, ou superior a 558.795, e nas escolas públicas esse número é de  $\approx 33\%$ .

Lucena e Dos Santos (2020) também constaram em seu estudo com os dados do Enem de 2016, que alunos que estudaram o ensino médio apenas em escolas públicas possuem 41,70 pts. a menos que os candidatos que estudaram todo o ensino médio em escolas particulares. E em nossa pesquisa, identificamos através das análises, que os candidatos do RJ que estudaram em escolas públicas tiveram a mediana de suas notas 97,48 pts menor que as notas dos

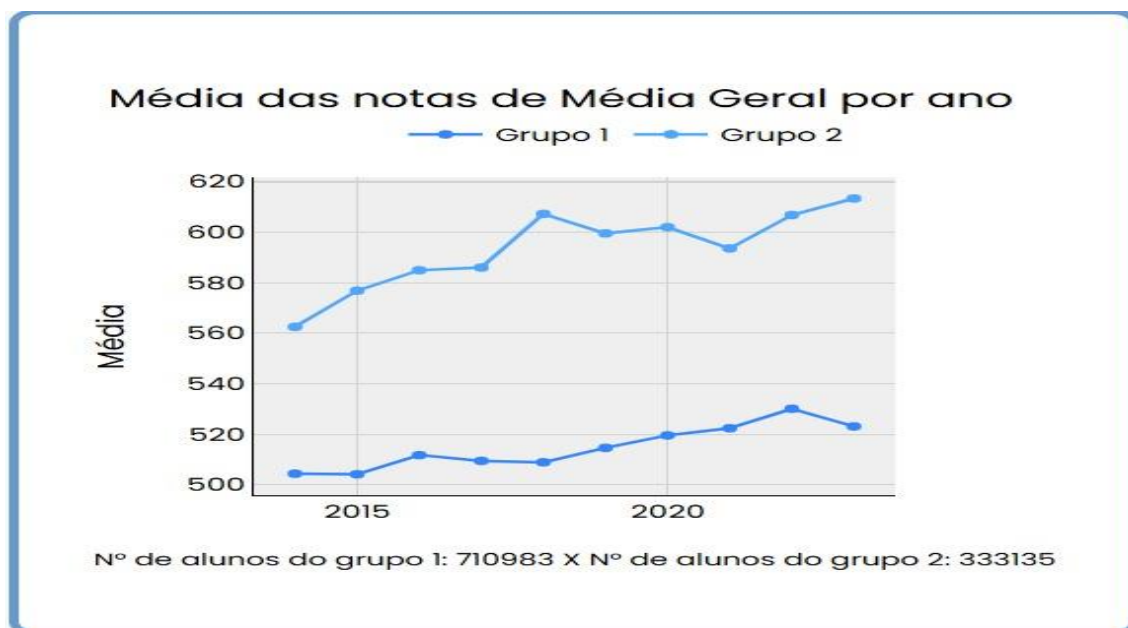
candidatos de escolas particulares, o que demonstra similaridade nos resultados quanto a hegemonia das escolas particulares nas melhores notas no Enem. Além disso, em nosso trabalho o *Upper Fence* (Limite Superior) coincidiu com a nota máxima nas privadas, o que sugere mais homogeneidade na distribuição das notas, enquanto nas públicas as notas máximas sendo *outliers* (valores discrepantes) sugere que alguns alunos se destacam por motivos desconhecidos.



**Figura 2** – Comparativos das notas da rede pública x privada em 2023.

Fonte: Autores (2025).

Entretanto, esse estudo expande o conhecimento gerado nos trabalhos anteriores, ao explorar uma série temporal mais ampla, de dez anos (2014-2023), figura 3. Rocha et al. (2022) explorou em sua pesquisa três anos, e Lucena; Dos Santos, (2020), abordaram apenas os dados de 2016. Essa maior abrangência de dados, possível através da ferramenta desenvolvida, permitiu identificar uma tendência crítica: a diferença no desempenho entre as escolas, que era de 58,12 pontos em 2014, saltou para 90,22 pontos em 2023, um aumento de 55% alcançado pelas privadas ao longo da série. O crescimento médio anual de notas na rede privada foi de +5,64 pontos, quase o triplo do crescimento de +2,08 pontos/ano observado na rede pública. Enquanto na privada houve um crescimento no período de 9%, a pública alcançou +3,7%. Esses dados não apenas reforçam a existência de uma desigualdade histórica, já registrada na literatura, mas também sugerem que as disparidades estão se acentuando ao longo do tempo, e que as políticas públicas implementadas não estão sendo capazes de combatê-las.



**Figura 3** – Evolução das notas, rede pública e rede privada de 2013 a 2023.

Fonte: Autores (2025).

### Cenário 2 (Renda Familiar)

Na amostra do cenário 2, figura 4, ficou confirmado a influência da renda familiar no desempenho dos estudantes, corroborando com achados dos estudos anteriores de Rocha et al. (2022), Lucena; Dos Santos, (2020), Santos (2019). Segundo Rocha et al. (2022) a média entre alunos de baixa renda na região nordeste entre 2017 a 2019 ficou menor ou igual a 500 pts. E Santos (2019), ao analisar os dados de 2018, constatou que os candidatos com renda  $\leq 2$  SM não conseguiram alcançar a média de pontos nacional, que foi de 538,40 pts. Nossa pesquisa corrobora com os achados de Rocha et al., pois em nossas análises identificamos que  $\approx 39\%$  dos candidatos com renda  $\leq 2$  SM não conseguiram alcançar 500 pts, a metade no exame, enquanto na amostra com renda  $\geq 10$  SM, esse valor foi de  $\approx 7\%$ , representando uma diferença expressiva entre os grupos analisados.

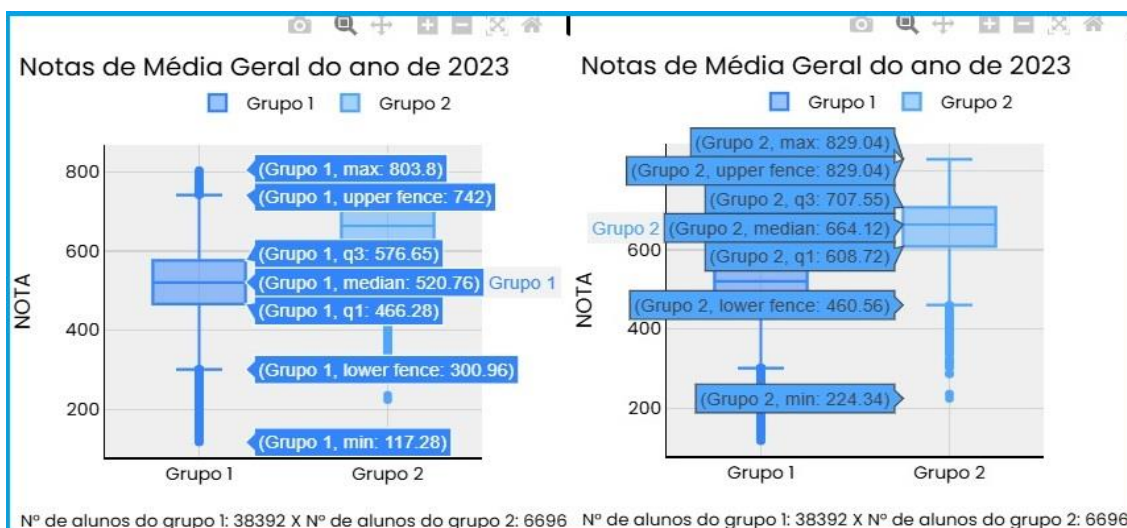
Já Lucena; Dos Santos, (2020), afirmam que os candidatos com renda  $\leq 2$  SM, no ano de 2016, possuem em média 83,49 pontos a menos que os candidatos de maior faixa de renda familiar. Nesse ponto, o presente estudo chegou a constatações semelhantes, pois identificamos que alunos do Grupo 1 do RJ, com renda menor ou igual a dois salários-mínimos ( $\leq 2$  SM), apresentam uma mediana de 520,76 pts, contra 664,12 pts do outro grupo, uma diferença de menos 143,36 pts (-27,52%) em relação ao Grupo 2, com renda maior ou igual a dez salários-mínimos ( $\geq 10$  SM). Essa diferença chama a atenção, pois se revela maior que a disparidade entre redes públicas e privadas (97,48 pts no Cenário 1), destacando que a renda familiar é um

preditor mais forte que o tipo de escola. O que também foi apontado por Travitzki et al. (2016), que ao analisar os microdados dos estados do Maranhão, Alagoas, Rio Grande do Sul e São Paulo, constatou que as disparidades na educação coincidem com as desigualdades na renda familiar.

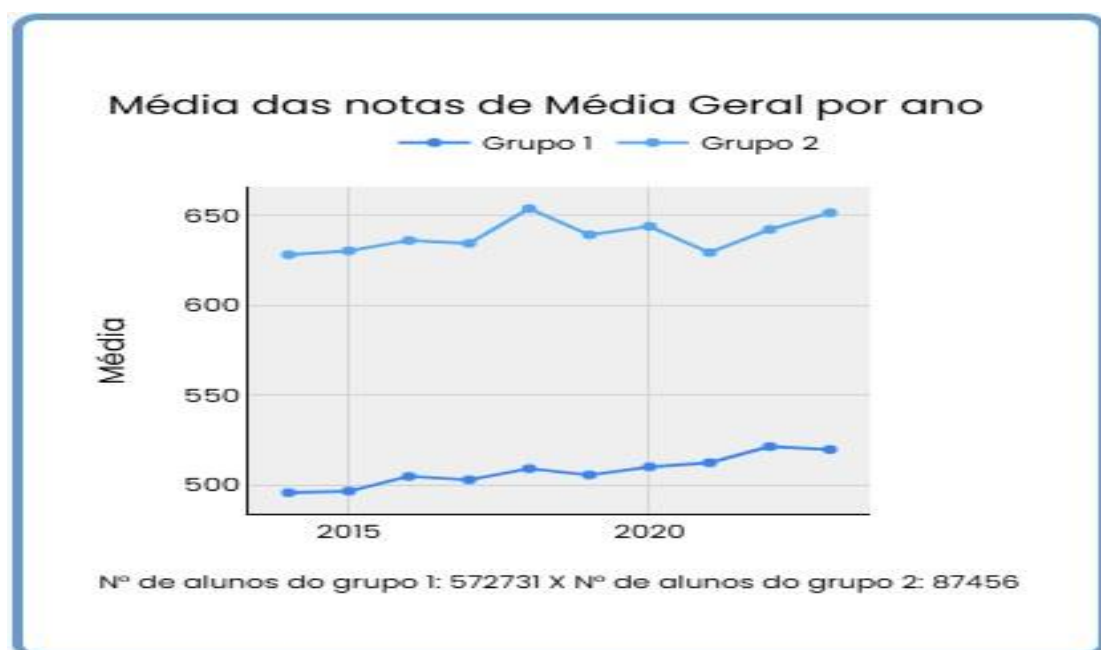
Além disso, nossa pesquisa identificou que 25% dos alunos de alta renda do RJ têm notas mais altas que 75% dos que possuem baixa renda, pois o quartil 1 (Q1) do Grupo 2, figura 4, possui um valor de 608,72 pts, que é superior a Q3 do Grupo 1, 576,65 pts, o que de certa forma já era esperado, como demonstra estudos anteriores, mais não com diferenças tão expressivas como encontrado. No grupo  $\geq 10$  SM a análise identificou que 75% dos alunos da amostra conseguiram notas  $\geq 608,72$  pts, nos alunos de renda baixa esse número é de  $\approx 21\%$ .

Este trabalho foi capaz de contribuir, e ampliar a literatura criada por estudos anteriores que analisaram as influências da renda nos resultados do Enem. Enquanto a pesquisa de Rocha et al. (2022) comprovou desigualdades motivadas por renda entre 2017 a 2019, Lucena; Dos Santos, (2020) em 2016, e Santos (2019) em 2018, este trabalho analisou a tendência temporal de dez anos, figura 5, e constatou a persistência do problema na última década. Revelando que a diferença entre grupos socioeconômicos distintos persistiu, mantendo-se instável e superando os 100 pts ao longo do período. Na pesquisa de Santos (2019) as análises identificaram que estudantes com renda familiar entre 5 a 10 salários-mínimos chegam a ter 67,54 pts a mais que os de renda entre 1 a 2 salários, aumentando para 105,99 pts entre os com renda maior que 10 salários. Em nosso estudo, o grupo com renda  $\geq 10$  SM manteve-se dominante, durante os 10 anos, atingindo seu ápice em 2018 com +145,06 pts (+28%) a frente do outro grupo, e obtendo uma média de  $\sim 132$  pts de diferença.

Ribeiro e De Oliveira (2024) argumentam que as disparidades socioeconômicas continuam a impor barreiras significativas ao desenvolvimento educacional equitativo. Sendo assim, nossa pesquisa não só comprovou o que já demonstrava a literatura, mas ampliou os resultados, mostrando que tais diferenças vem se perpetuando a pelo menos uma década. Isso reforça a ideia de que a desigualdade socioeconômica é um padrão enraizado e resistente a mudanças, servindo como um forte argumento para a necessidade de novas abordagens.



**Figura 4** – Comparação entre renda familiar-  
Fonte: Autores (2025).



**Figura 5** – Comparativo da Evolução das notas, baseado na renda familiar-  
Fonte: Autores (2025).

### Cenário 3 (Etnia Em Estados Contrastantes)

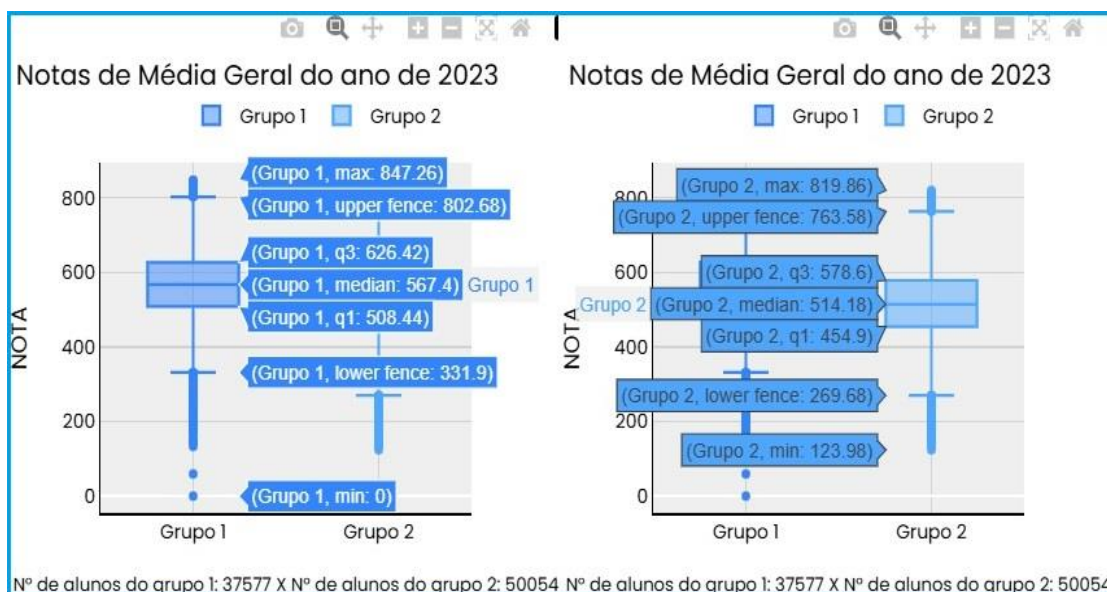
As análises realizadas no Cenário 3, reforçaram a tese de que a justiça social como reconhecimento transita da igualdade formal para a material, demonstrando que as disparidades educacionais também estão profundamente enraizadas em fatores como etnia e região. Essa constatação corrobora com os achados de Ribeiro e De Oliveira (2024), que destacam as persistentes desigualdades regionais e socioeconômicas no acesso à educação de qualidade. Ressaltando em seu estudo que as regiões Norte e Nordeste apresentam resultados inferiores em comparação com o Sul e Sudeste.

Santos, 2019, em suas análises constatou que as regiões Sudeste e Sul foram as com maiores médias nacionais no ano de 2018, 556 e 548 pts respectivamente, enquanto as regiões Norte e Nordeste, 512 e 523 pts, as menores. Nesse ponto, a ferramenta desenvolvida confirmou os resultados de Santos, ao identificar na análise da distribuição das notas que 42% dos alunos da Bahia obtiveram notas abaixo de 500, contra apenas 24% no Rio Grande do Sul. E que 78% dos alunos da BA possuem notas inferiores a 600 pts, sendo no outro grupo 62%, demonstrando que a região Sul possui maior desempenho do que a região Nordeste do país, reforçando o argumento de Ribeiro e De Oliveira (2024) que desigualdades regionais no acesso à educação de qualidade é uma realidade presente.

Lucena; Dos Santos, (2020), ao responderem a hipóteses 6 (H 6) “Os candidatos de cor branca tendem a ter desempenho melhor na nota geral do Enem”, afirmam que os candidatos de cor branca tendem a ter melhores resultados na nota geral que os demais candidatos pretos, pardos, amarelos e indígenas. Nosso estudo enriquece esse debate ao comparar estados historicamente contrastantes em sua composição socioeconômica e étnica, como RS e BA, pois como resultado tivemos alunos brancos do Rio Grande do Sul, Grupo 1, com uma mediana de +53,22 pts (10,3%) em relação ao Grupo 2 (BA) formado por alunos pretos e pardos, figura 6, o que confirma as disparidades regionais e étnicas do Brasil, apontadas em outros estudos.

Nosso estudo expande a discussão ao conectar a teoria de “justiça social como redistribuição e reconhecimento” com evidências baseadas em dados reais. O estudo de Ribeiro e De Oliveira (2024) propõe que as políticas de equidade devem reconhecer as disparidades de raça e região, e o presente estudo, com a comparação entre grupos étnicos em estados contrastantes, propõe exatamente isso, que justiça social e distributiva (equidade), devem ser orientados pelo critério socioeconômico, que corresponde ao ideal de justiça como reconhecimento de raça, idade, gênero, etnia e demais critérios.

Além disso, nossa análise de dez anos aprofunda ainda mais os debates sobre o tema, pois mostraram que a diferença entre os Grupos 1 e 2, aumentou de 37,73 pts em 2014, para 48,73 pts em 2023, um aumento de 29%, tendo o Grupo 1 (RS) predominado em todo o período, conforme figura 7. Os resultados das análises indicam que o problema não apenas persiste, mas se agrava com o tempo, esse agravamento das disparidades ao longo do tempo reforça a necessidade de políticas educacionais que reconheçam as especificidades regionais e étnicas para garantir um avanço real em direção à equidade.



**Figura 6** – Comparativo baseado na etnia dos estados contrastantes (2023).  
 Fonte: Autores (2025).



**Figura 7** – Evolução das notas, baseado na etnia (2014 a 2023).  
 Fonte: Autores (2025).

### Contribuição Metodológica, Limitações e Trabalhos Futuros

A principal contribuição metodológica deste estudo é o desenvolvimento de uma aplicação interativa que oferece uma nova abordagem para a análise de grandes volumes de dados educacionais. A ferramenta permite a investigação de padrões de desigualdade em um lapso temporal de dez anos, preenchendo lacunas metodológicas identificadas em pesquisas anteriores, como a limitação temporal das bases de dados que normalmente são de 1-3 anos, e

ainda superando o desafio da restrição geográfica (análise de poucos estados), já que a ferramenta desenvolvida permite que todos os estados brasileiros sejam analisados.

Mas apesar da sua relevância, a pesquisa apresenta algumas limitações. A aplicação, embora tenha sido desenvolvida com o intuito de se tornar pública, não foi hospedada em um servidor devido à falta de recursos financeiros. Por esse motivo, não foi possível realizar testes de usabilidade com usuários finais externos, o que seria crucial para o aprimoramento de suas funções e intuitividade. Em termos de escopo, a aplicação limitou-se a consultas por estados, sem permitir análises por regiões específicas do país, e utilizou apenas uma fração das variáveis disponíveis nos microdados do Enem, devido à complexidade que mais opções trariam a seu desenvolvimento. Além disso, a natureza quantitativa dos microdados não permite captar fatores qualitativos que podem vir a influenciar na performance acadêmica, como a metodologia pedagógica, qualificação dos docentes, ou os efeitos da pandemia.

Diante dessas limitações, é possível sugerir uma série de caminhos para pesquisas futuras. Acreditamos que a principal delas seria a disponibilização da ferramenta em um servidor para torná-la acessível ao público, permitindo que pesquisadores, educadores e formuladores de políticas públicas possam realizar suas próprias análises. A realização de testes de usabilidade também é fundamental para validar a ferramenta e aprimorá-la. Aprimoramentos futuros poderiam incluir consultas por regiões do Brasil, exploração de outras variáveis presentes nos microdados do Enem, e ainda a integração com dados qualitativos para uma compreensão mais completa das desigualdades educacionais no país, são algumas das possibilidades para desenvolvimento de trabalhos futuros.

### **Considerações Finais**

As considerações finais deste trabalho apontam para uma reflexão profunda sobre as desigualdades no sistema educacional brasileiro, analisadas sob a lente da equidade educacional e observadas a partir de três dimensões fundamentais: o tipo de escola, a renda familiar e a interseccionalidade entre etnia e região. Integrar uma década de dados do Enem e abranger todos os estados em uma aplicação interativa não foi apenas um exercício metodológico: foi, sobretudo, um convite a enxergar aquilo que muitas vezes permanece invisível sob o peso das estatísticas.

As hipóteses que guiaram esta investigação confirmaram-se, mas os resultados vão além da confirmação. A disparidade entre escolas públicas e privadas não só persiste como se ampliou em mais de 50% ao longo de dez anos, revelando que a promessa de igualdade de oportunidades ainda não encontrou terreno fértil. A renda, como um fio condutor inevitável,

mostrou-se o preditor mais poderoso, lembrando-nos de que a educação, no Brasil, continua refém das condições econômicas que moldam o destino de famílias inteiras. Já o cruzamento entre etnia e região expôs algo ainda mais complexo: o Rio Grande do Sul manteve-se à frente da Bahia em todos os anos analisados, o que remete à relação direta entre desenvolvimento econômico regional e oportunidades educacionais. Apesar de a Bahia possuir o sétimo maior PIB do país, a diferença de 47% em relação ao PIB do Rio Grande do Sul, associada à má distribuição de renda, reforça a existência de um “teto de oportunidades” inalcançável para a maioria de baixa renda. A desigualdade racial não se limita ao indivíduo, mas é amplificada pelos territórios. A constância com que alunos brancos do Rio Grande do Sul superaram, ano após ano, os negros e pardos da Bahia não revela apenas diferenças de desempenho, mas escancara o quanto a geografia e a economia do país também determinam quem pode sonhar mais alto.

Quanto a aplicação desenvolvida para este estudo foi, nesse sentido, mais do que uma ferramenta tecnológica: revelou padrões estruturais que ultrapassam as meras barreiras das diferenças entre indivíduos, pois apontam para disparidades enraizadas no sistema de educação nacional. Demonstrou ser eficaz na produção e visualização de dados, baseando-se em variáveis socioeconômicas e sociodemográficas, contribuindo cientificamente para a literatura sobre equidade educacional. Além disso, a ferramenta também apresenta um tempo de resposta aceitável, seis a oito segundos, diante do volume de dados, e uma interface intuitiva, alcançando os objetivos da sua proposta de criação.

Com isso, o sistema elaborado cumpriu um papel fundamental ao inovar na integração e visualização dos dados, e possibilitou, não apenas análises de grande escala, mas também uma exploração mais intuitiva e interativa, tornando viável a realização de análises exploratórias de dados (AED) em um nível até então pouco acessível em estudos educacionais. Mostrou-se eficiente ao associar desempenho escolar a variáveis socioeconômicas e demográficas, produzindo resultados de relevância científica e apresentando um modelo replicável para pesquisas futuras.

### **Referências:**

- Aldridge, J. M., & Fraser, B. J. (2008). *Outcomes-focused learning environments*. Brill.
- Barbosa, E. S., Cruz, M. C. M. T., & Ribeiro, V. M. (2022). Educational equity in vulnerable territories. *Revista Tempos e Espaços em Educação*, 15(34), e16958. <https://doi.org/10.20952/revtee.v15i34.16958>
- Brasil. (1988). *Constituição da República Federativa do Brasil*. Senado Federal.

- Brasil. (1996). *Lei n. 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional*. [http://www.planalto.gov.br/ccivil\\_03/leis/19394.htm](http://www.planalto.gov.br/ccivil_03/leis/19394.htm)
- Brasil. (2001). *Lei n. 10.172, de 9 de janeiro de 2001. Aprova o Plano Nacional de Educação - PNE e dá outras providências*. [http://www.planalto.gov.br/ccivil\\_03/leis/leis\\_2001/110172.htm](http://www.planalto.gov.br/ccivil_03/leis/leis_2001/110172.htm)
- Brasil. (2014). *Lei n. 13.005, de 25 de junho de 2014. Aprova o Plano Nacional de Educação - PNE e dá outras providências*. [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2014/lei/113005.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/113005.htm)
- Brasil. Governo Federal – Serviço e Informações do Brasil. (2025). *Inscritos confirmados no Enem*.
- Crahay, M. (2000). *Poderá a escola ser justa e eficaz? Da igualdade das oportunidades à igualdade dos conhecimentos*. Instituto Piaget.
- Dos Santos, J. A., Silva, L. M., Oliveira, R. C., & Pereira, M. A. (2024). Desafios e oportunidades do ensino em computação: Uma análise exploratória e multivariada dos dados do ENADE. *Revista Intersaberes*, 2024, e24tl4019. <https://doi.org/10.22169/revint.v19.e24tl4019>
- Dubet, F. (2008). Democratização escolar e justiça da escola. *Educação*, 33(3), 381–394. <https://periodicos.ufsm.br/reeducacao/article/view/1614>
- Ferreira, L., Oliveira, F. S., Rocha, O. P., Holanda, M., Carvalho Victorino, M., & Ribeiro, E. (2021). MongoDB: Analysis of performance with data from the national high school exam (ENEM). In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–6). IEEE.
- Futura. (2023). *PISA 2022: Por que o Brasil está nas últimas posições em matemática, ciências e leitura?*
- INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2022). *Divulgados os resultados do PISA 2022*.
- Lopes, G. R., Almeida, A. W. S., Delbem, A. C. B., & Toledo, C. F. M. (2019). Introdução à análise exploratória de dados com Python. *Minicursos ERCAS ENUCMPI, 2019*, 160–176.
- Lucena, J. P. O., & Dos Santos, H. N. L. (2020). A relação entre desempenho no Exame Nacional do Ensino Médio e o perfil socioeconômico: Um estudo com os microdados de 2016. *Revista de Gestão e Secretariado*, 11(2), 1–23. <https://doi.org/10.7769/gesec.v11i2.994>
- Microsoft. (2024). *Arquivo Apache Parquet*.
- Nakazone, E., & Bortolotti, L. M. (2021). Análise de dados históricos do ENEM entre 2015 a 2019. In *Anais do Congresso de Tecnologia – Fatec Mococa*.
- Piovesan, F. (2005). Ações afirmativas da perspectiva dos direitos humanos. *Cadernos de Pesquisa*, 35(124), 43–55. <https://doi.org/10.1590/S0100-15742005000100004>
- Portal Insights. (2024). *Qual a origem da palavra equidade?*
- Prodanov, C. C., & Freitas, E. C. (2013). *Metodologia do trabalho científico: Métodos e técnicas da pesquisa e do trabalho acadêmico* (2. ed.). Editora Feevale.
- Ribeiro, M. C., & Oliveira, R. R. F. de. (2024). Entre avanços e desafios: O caminho da educação brasileira rumo à equidade. *Observatório de a Economía Latino-Americana*, 22(5), e4650. <https://doi.org/10.55905/oelv22n5-097>

- Ribeiro, V. M. (2014). Que princípio de justiça para a educação básica? *Cadernos de Pesquisa*, 44(154), 1094–1109. <https://doi.org/10.1590/19805314-2844>
- Rocha, F. B. N., Silva, A. M., Costa, P. R., & Lima, R. S. (2022). Análise de microdados de matemática do ENEM 2017-2019 do Nordeste. *Pesquisa, Sociedade e Desenvolvimento*, 11(10), e111032716. <https://doi.org/10.33448/rsd-v11i10.32716>
- Rodrigues, D. D. C., Santos, M. A., Oliveira, L. P., & Fernandes, R. C. (2019). A data mining approach applied to the high school national examination: Analysis of aspects of candidates to Brazilian universities. In *Progress in Artificial Intelligence* (pp. 3–14). Springer. [https://doi.org/10.1007/978-3-030-30241-2\\_1](https://doi.org/10.1007/978-3-030-30241-2_1)
- Sampaio, G. T. C., & De Oliveira, R. L. P. (2015). Dimensões da desigualdade educacional no Brasil. *Revista Brasileira de Política e Administração da Educação*, 31(3), 511–530. <https://doi.org/10.21573/vol31n32015.60121>
- Santos, G. Q. (2019). *Os efeitos das desigualdades regionais nos resultados do Enem: Uma análise a partir dos microdados de 2018* [Trabalho de Conclusão de Curso, Universidade Federal do Rio Grande do Sul]. Lume Repositório Digital.
- Soares, J. F., & Delgado, V. M. S. (2016). Medida das desigualdades de aprendizado entre estudantes de ensino fundamental. *Estudos em Avaliação Educacional*, 27(66), 754–780. <https://doi.org/10.18222/ae.v27i66.4101>
- Travitzki, R., Ferrão, M. E., & Couto, A. P. (2016). Desigualdades educacionais e socioeconômicas na população brasileira pré-universitária: Uma visão a partir da análise de dados do ENEM. *Education Policy Analysis Archives*, 24, 1–32. <https://doi.org/10.14507/epaa.24.2199>
- Vicentainer, D., Mattedi, M., & Mello, B. (2020). Aplicação das bibliotecas Python para tratamento de dados em tempo real: A análise dos dados de isolamento social em Santa Catarina. *Metodologias e Aprendizado*, 3, 206–217.
- Villani, M., & Oliveira, D. A. (2018). Avaliação Nacional e Internacional no Brasil: Os vínculos entre o PISA e o IDEB. *Educação & Realidade*, 43(4), 1343–1362. <https://doi.org/10.1590/2175-62368-4893>

Submetido em 15/07/2025

Aceito em 25/09/2025