



Uma Análise do Desempenho dos Estudantes do Rio Grande do Sul no ENEM 2019

Rafael Vinícios do Carmo – Universidade Feevale – rafaelvcarmo@gmail.com
 Wesllel Felipe Heckler – Universidade Feevale – weslleiheckler@gmail.com
 Juliano Varella de Carvalho – Universidade Feevale – julianovc@feevale.br

Resumo. *O Exame Nacional do Ensino Médio (ENEM) avalia o desempenho de estudantes do ensino médio do país. Este estudo explora os resultados do ENEM 2019, com foco nos dados do estado do Rio Grande do Sul. O estudo analisa e compara o perfil educacional e socioeconômico dos participantes que se enquadram nas 5% melhores e piores médias do estado. Os resultados demonstram uma tendência de melhor desempenho para participantes das classes sociais A e B; oriundos de escolas privadas e federais; que são brancos, pardos ou amarelos; que possuem acesso à internet em casa e cujos pais possuem maior grau de ensino. Essa análise evidencia os impactos das diferenças socioeconômicas no desempenho dos alunos no exame.*

Palavras-chave: ENEM, Desempenho de Estudantes, Perfil de Estudantes, Análise de Dados, Extração de Conhecimento

An Analysis of the Performance of Students from the Rio Grande do Sul in ENEM 2019

Abstract. *National Exam of High School (ENEM) evaluates the performance of the High School students from Brazil. This study explores the results of ENEM 2019, focusing on the data of students from the Rio Grande do Sul state. The study analyzes and compares the educational and socioeconomic profile of participants who are in the groups of 5% best and worst average results of the state. The results show a tendency for better performance for those students that are from the social classes A and B; who arise from private and federal schools; who are white, brown, or yellow; who have internet access at home; and whose parents have a superior degree of study. This analysis demonstrates the impacts of socioeconomic differences in students performance in the exam.*

Keywords: ENEM, Students Performance, Students Profile, Data Analysis, Knowledge Extraction

1. Introdução

O Exame Nacional do Ensino Médio (ENEM) é aplicado anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) para medir a qualidade do ensino médio brasileiro. Ele também colabora para o acesso à educação superior, servindo como base para a distribuição de bolsas e oportunidades de financiamento. O exame é composto por quatro provas objetivas referentes às áreas Linguagens e códigos, Ciências humanas e suas tecnologias, Ciências da natureza e suas tecnologias e Matemática e suas tecnologias e uma redação dissertativa-argumentativa (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2020).

A primeira edição do exame foi realizada em 1998 e reuniu 157,2 mil inscritos e 115,6 mil participantes. Um grande aumento de participantes ocorreu em 2004 quando o



Programa Universidade para Todos (ProUni) foi criado pelo Ministério da Educação, que também passou a utilizar o desempenho na prova para distribuir bolsas em Instituições de Ensino Superior privadas (Instituto Brasileiro de Geografia e Estatística, 2020).

Em 2007, o governo criou o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI), que visa ampliar o acesso e permanência no ensino superior. O REUNI foi responsável por expandir o número de universidades no Brasil, criando diversos campi no interior dos estados (de Paula, 2017). O Programa contribuiu de forma significativa para o aumento de participantes das regiões Norte e Nordeste do Brasil a partir de 2010, tornando a proporção de participantes por região mais próxima à proporção populacional das regiões. O número de inscritos aumentou consideravelmente ano após ano até 2014, atingindo quase 9 milhões de estudantes, quando começou a sofrer oscilações e diminuir (Santos et al., 2019). Em 2019, o número de inscritos foi de 5.095.270.

O INEP armazena informações sobre cada estudante inscrito no exame, formando uma grande base de dados, que é disponibilizada para cada edição realizada. Entre as informações armazenadas, estão dados sobre os estudantes, tais como idade, sexo, residência, entre outros. Também são armazenadas informações sobre a escola onde o inscrito estuda, o desempenho nas provas objetivas e na redação e as respostas para um questionário que mapeia o perfil socioeconômico de cada estudante. A quantidade de informações registradas sobre os participantes somada ao amplo número de inscritos torna a base de dados do ENEM uma rica fonte de informações, que pode ser explorada, a fim de identificar informações relevantes sobre o ensino brasileiro.

Segundo Han et al. (2012), mineração de dados é o processo de descoberta de conhecimento sobre uma grande quantidade de dados, que podem estar armazenados em diversas fontes. Através dela, os padrões podem ser extraídos, analisados e apresentados visualmente para gerar conhecimento significativo sobre os dados. A visualização de dados visa tornar mais clara e efetiva a comunicação com a utilização de elementos visuais. É possível utilizar as técnicas de visualização para descobrir relações de dados que, de outra forma, não seriam facilmente encontradas. Assim, a visualização de dados facilita o entendimento humano sobre determinado conjunto de dados, melhorando a tomada de decisões e a formação de hipóteses (Webber et al., 2014).

A mineração de dados abrange aspectos de estatística, tecnologia da informação, inteligência artificial e visualização de dados. Essa tecnologia facilita a extração de conhecimento e a análise sobre os dados, tarefas que demandam uma grande quantidade de tempo se realizadas manualmente. A mineração de dados educacionais é um campo que está em crescimento atualmente (Salloum et al., 2020). Ela fornece um melhor entendimento sobre o comportamento e estilo de aprendizagem dos estudantes. Assim, pode-se analisar aspectos do sistema educacional com o objetivo de melhorar a qualidade da educação através da criação de políticas educacionais (Fernandes et al., 2019).

Segundo Baker (2014), a mineração de dados educacionais é dividida em quatro principais áreas, sendo elas aprimorar o desempenho dos estudantes, descobrir ou melhorar modelos de conhecimento de domínio (por exemplo, predição de ações de estudantes), estudar o suporte pedagógico através de sistemas de aprendizagem e, por fim, aplicar a mineração de dados para descobertas científicas sobre a aprendizagem e os



alunos. O advento de repositórios públicos de dados educacionais torna viável essa tarefa. A existência de dados sobre milhares de estudantes possibilita a análise da influência de fatores contextuais no ensino e nos estudantes. A mineração de dados educacionais tem o potencial de fornecer um conjunto de ferramentas para análise de questões importantes, considerando as diferenças individuais e facilitando uma análise que, historicamente, é complexa.

Este trabalho, portanto, apresenta um estudo sobre a base de dados do ENEM 2019, com foco nos estudantes do Rio Grande do Sul. Através de visualizações sobre os dados, são analisados padrões sobre o perfil dos estudantes que estão entre os 5% melhores colocados do estado no exame em comparação com os que se enquadram no grupo de 5% piores colocados, de acordo com a média aritmética geral. Entre os atributos utilizados para esta análise, estão o acesso a recursos digitais, como, por exemplo, à internet; renda familiar; sexo; tipo de escola; cor/raça; e grau de estudo dos pais.

Este artigo está estruturado em 5 seções. A seção 2 descreve a metodologia utilizada neste trabalho. A seção 3 discute trabalhos relacionados. Os resultados são discutidos na seção 4. Por fim, a seção 5 apresenta as conclusões deste trabalho.

2. Materiais e Métodos

Inicialmente foi realizado um estudo sobre os metadados da base de dados do ENEM, que é disponibilizada anualmente pelo INEP*. As 136 variáveis presentes no *dataset* foram avaliadas. A análise inicial teve como objetivo identificar relações entre os dados do perfil dos estudantes e seus respectivos desempenhos. Através desse estudo, foram selecionados os atributos de interesse para geração de visualizações.

Posteriormente foram investigadas e selecionadas técnicas de apresentação gráfica que permitissem a representação de padrões sobre o perfil educacional e socioeconômico dos participantes, considerando os atributos selecionados no estudo inicial, tais como grau de instrução dos pais, acesso à internet, renda familiar, dentre outras variáveis. Após isso, foi realizada uma análise sobre os municípios do Rio Grande do Sul, visando possibilitar a exploração de características de desempenho dos municípios do estado.

O estudo sobre os dados e a geração das visualizações foram realizados com a linguagem de programação Python[†] e a ferramenta Jupyter Notebook[‡]. Para a geração de visualizações baseadas em mapas do território brasileiro, utilizou-se a *Application Programming Interface* (API) que disponibiliza as coordenadas do país, estados e municípios, através de requisições. Esse serviço de dados é disponibilizado pelo IBGE[§].

3. Trabalhos Relacionados

Esta seção apresenta trabalhos relacionados ao tema de pesquisa. Esses trabalhos servem como base para análise dos resultados encontrados neste artigo. Existem poucos trabalhos que visam estudar o perfil dos estudantes do exame. Em geral, os trabalhos avaliam o desempenho dos estudantes, considerando todo o território brasileiro e não apenas um

*<http://inep.gov.br/microdados>

†<https://www.python.org/>

‡<https://jupyter.org/>

§<https://servicodados.ibge.gov.br/api/docs/malhas?versao=2>



estado. Também não foi encontrado um trabalho que analisou as características dos alunos com base nos grupos de melhores e piores médias.

O trabalho de Santos et al. (2019) analisou os dados dos exames realizados entre 1998 e 2017, a fim de avaliar as mudanças no perfil dos estudantes nesse período. A pesquisa consistiu em analisar características sociais, geográficas e econômicas, tais como sexo, idade, raça, região, classe social e grau de estudo dos pais, apresentando as variações de cada característica ao longo das 20 edições analisadas. Zacchi et al. (2016) avaliaram os dados do ENEM 2011 com o objetivo de identificar a relação entre os recursos socioeconômicos familiares e o desempenho dos alunos. Esta pesquisa considerou os alunos entre 17 e 19 anos de idade, que informaram a dependência administrativa da escola e que completaram todas as quatro provas objetivas.

Martins et al. (2017) desenvolveram modelos preditivos para identificar desistências de alunos em cursos de ensino superior. A evasão de alunos é um problema para as instituições. Esse problema se agrava quando envolve as universidades públicas, visto que as desistências geram desperdício de recursos públicos nesse contexto, além de grande prejuízo ao desenvolvimento educacional dos estudantes. O trabalho também tem como objetivo identificar o perfil dos estudantes que abandonam os cursos superiores. Desta forma, é possível implementar medidas para evitar a evasão.

Viggiano e Mattos (2013) utilizam os dados do ENEM 2010 para avaliar o desempenho de estudantes. Os autores comparam os resultados dos estudantes de todas as regiões geográficas do Brasil em cada uma das áreas de conhecimento avaliadas pelo exame. Corti (2013) analisou os dados do exame nas edições de 1999 a 2007, fazendo reflexões sobre o perfil dos participantes de todas as regiões brasileiras.

4. Resultados

O *dataset* do ENEM armazena uma grande quantidade de dados. A análise desses dados revela padrões sobre a qualidade do ensino e sobre o desempenho dos estudantes brasileiros. O registro de dados geográficos permite uma análise mais detalhada com foco nos municípios, estados e regiões do Brasil. Assim, é possível considerar na análise as diferenças culturais e socioeconômicas existentes ao longo do território brasileiro.

Desta forma, esta seção descreve a análise realizada sobre os dados do ENEM 2019 com foco nos estudantes do Rio Grande do Sul. O objetivo da análise é a identificação de padrões sobre o desempenho dos municípios e estudantes gaúchos no exame. Através de visualizações, os resultados dos municípios são analisados. Além disso, são analisados os padrões socioeconômicos dos estudantes que se enquadram nos grupos de 5% melhores e piores colocados. Essa classificação foi baseada na média aritmética das notas de cada aluno, considerando as quatro provas objetivas (Ciências da Natureza, Ciências Humanas, Linguagens e Códigos e Matemática) e a redação.

4.1. Preparação dos Dados

O *dataset* do ENEM 2019 possui 5.095.270 registros, que representam os estudantes, e 136 atributos, que são as informações armazenadas para cada estudante. Os dados estão divididos em 9 seções. A primeira seção apresenta dados sobre o participante, tais como sexo, idade e cor/raça. A segunda apresenta dados da escola, como localização, tipo de dependência administrativa e situação de funcionamento. A terceira, quarta e quinta



seções apresentam dados sobre atendimento especializado para estudantes com deficiência, gestantes e idosos. Dados sobre o local de aplicação da prova são apresentados na sexta seção. A sétima e oitava seções expõem dados sobre as provas objetivas e a redação. A nona seção, por fim, apresenta as respostas para um questionário socioeconômico, que mapeia o perfil dos estudantes, desde a renda familiar mensal até o grau de estudo dos pais.

Para a análise de desempenho, foram considerados apenas os registros de estudantes que estudam no Rio Grande do Sul. Estudantes que possuem, pelo menos, uma nota não registrada foram desconsiderados. Ou seja, estudantes que não realizaram alguma das provas por algum motivo. Ao todo, 39.055 registros foram considerados neste trabalho. Dentre os 136 atributos do *dataset* original, foram selecionados para análise o nome do município da escola, a Unidade Federativa da escola, as notas das provas objetivas, a nota da redação e as respostas para o questionário socioeconômico.

4.2. Análise de Desempenho dos Estudantes

A análise inicial consistiu em uma avaliação de resultados dos estados brasileiros, a fim de obter uma visão geral sobre o desempenho nas diferentes regiões brasileiras. A Figura 1 demonstra a média geral por estado. Nesta visualização, os estados das regiões Sul e Sudeste recebem destaque em relação aos estados das regiões Norte e Nordeste. Resultados semelhantes já haviam sido encontrados na pesquisa de Viggiano e Mattos (2013), que analisou o desempenho dos estudantes no ENEM 2010. O resultado dessa pesquisa mostra que a região Sudeste obteve as melhores médias, comparando com as demais regiões. Além disso, foi identificado que a região Norte teve as piores médias. Mesmo após quase 10 anos, ainda existe uma grande diferença de resultados entre as regiões. O Rio Grande do Sul se enquadra no grupo de estados que atingiu melhor desempenho entre os estados brasileiros, sendo o foco de análise deste artigo.

Para a visualização em mapa, a escala de cores representa a média geral, quanto mais escuro, maior a média. Os espaços em branco, na Figura 2, representam os municípios que não possuem dados para o cálculo de média geral. A Figura 2 apresenta o desempenho dos municípios do estado no exame. Essa visualização apresenta uma distribuição mais uniforme em relação à Figura 1. As cidades localizadas próximas às regiões Norte e Leste se destacam, assim como algumas da região central do estado. Os cinco municípios com melhores desempenhos são: Westfália, Coronel Pilar, Vila Lângaro, São Domingos do Sul e Mato Leitão. Já entre os piores desempenhos, estão: Novo Cabrais, Dois Irmãos das Missões, Nicolau Vergueiro, Inhacorá e Bom Progresso.

A partir da análise inicial, buscou-se identificar as características dos estudantes que se enquadram no grupo de 5% melhores médias gerais do Rio Grande do Sul, bem como as dos estudantes que se enquadram no grupo de 5% piores médias. Para a definição dos grupos, foi considerado o total de 39.055 estudantes, sendo essa a quantidade de estudantes que realizaram todas as provas. Cada grupo foi formado por 1.952 estudantes.

Nas Figuras 3 e 4 são apresentadas as distribuições de estudantes por raça em cada grupo analisado. Salienta-se que, o termo raça é usado pelo INEP. As raças estão ordenadas crescentemente, de acordo com a proporção de estudantes presentes em cada grupo sobre a quantidade total de participantes de cada raça no exame. Pode-se observar uma maior presença de estudantes das raças branca, indígena e amarela no grupo de

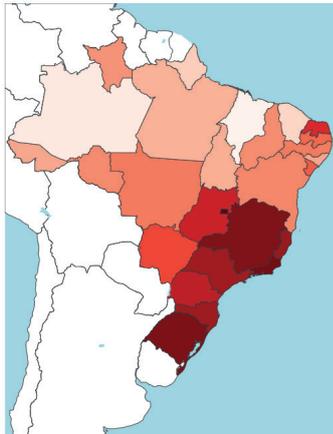


Figura 1. Mapa de desempenho dos estados

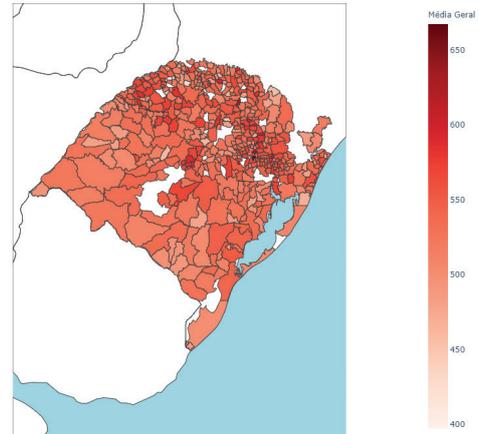


Figura 2. Mapa de desempenho dos municípios

estudantes que alcançou os melhores resultados. As raças preta e parda são as menos frequentes nesse grupo. Entre os estudantes que atingiram resultados piores, a raça branca é a menos frequente apesar de possuir maior quantidade geral de participantes. Nesse grupo, os estudantes das raças indígena e preta possuem maior proporção em relação às demais raças. A raça indígena recebe destaque nessa análise, pois aparece entre as mais frequentes nos dois grupos analisados. Destaca-se que, em todos os gráficos de barra, a legenda *Total* refere-se ao total de estudantes considerados na análise.

O exame possibilita que os estudantes não declarem suas raças, sendo esses classificados com raça não declarada, podendo interferir nos resultados analisados. Em 2013, Corti (2013) analisou o perfil dos estudantes sobre os exames realizados entre os anos de 2001 a 2007. Ele afirma que, ao longo dos 7 anos analisados, houve um aumento de 33% na presença de negros, pardos e mulatos na avaliação. O número de estudantes dessas raças, no entanto, segue sendo bastante baixo em relação às demais raças.

A pesquisa realizada por Zacchi et al. (2016) também identificou tendências de melhores desempenhos para alunos brancos em comparação aos pretos, considerando estudantes de todo o Brasil. Os estudantes brancos apresentam, em média, desempenho 8,2% superior aos pretos. Esta diferença, porém, é maior em alguns estados, como Rio de Janeiro, Sergipe, Espírito Santo e no Distrito Federal. No Rio de Janeiro, o desempenho de estudantes pretos chega a ser 10,6% menor.

As Figuras 5 e 6 demonstram a organização das classes sociais entre os grupos de melhores e piores desempenhos. Estudantes com classe social A e B constituem a maior parcela do grupo de melhores desempenhos apesar de representarem menor parcela na quantidade total de alunos que realizaram o exame. Já no outro grupo, estudantes com classe social D e E são os que constituem a maior parcela. Eles também representam a maior parcela de estudantes que realizaram o exame. Esse padrão evidencia os impactos socioeconômicos sobre o desempenho dos estudantes no ENEM.

Zacchi et al. (2016) identificaram resultados semelhantes, analisando dados do ENEM de 2011. A pesquisa demonstrou que, em todos os estados, o desempenho tende a aumentar conforme o aumento na renda familiar. Os autores também destacam a relação



Quantidade de estudantes por raça

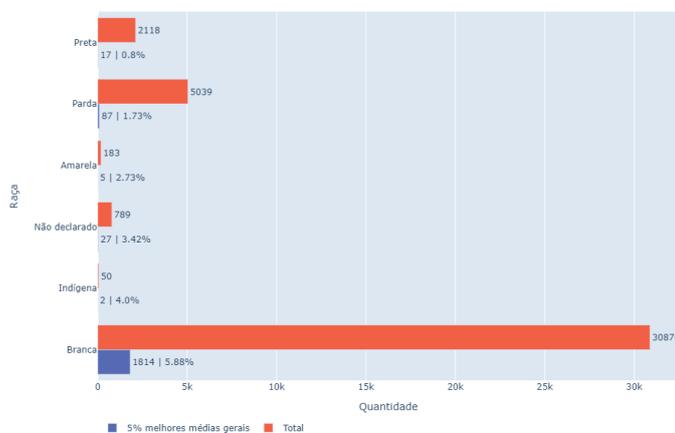


Figura 3. Melhores médias por raça

Quantidade de estudantes por raça

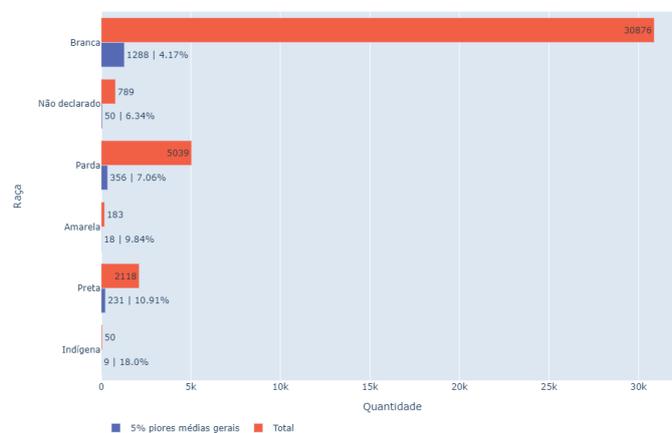


Figura 4. Piores médias por raça

Quantidade de estudantes por classe social

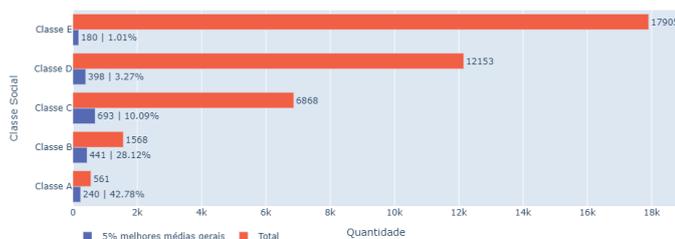


Figura 5. Melhores médias por classe social

Quantidade de estudantes por classe social

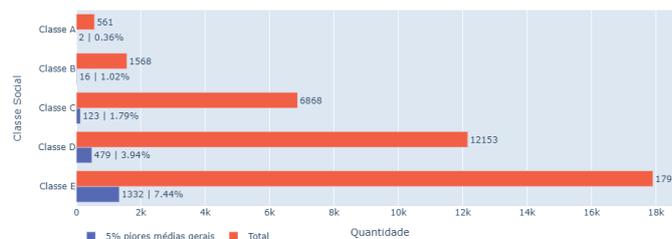


Figura 6. Piores médias por classe social

direta entre a renda familiar e o desempenho educacional no Brasil. Quanto menor a renda, menor tende a ser o investimento em educação, bem como o acesso à cultura, bens e serviços, tais como livros, viagens, cinemas, teatros, museus, entre outros.

As Figuras 7 e 8 exibem a disposição de estudantes em cada grupo conforme o grau de estudo da mãe. É possível observar que os alunos cuja mãe possui um grau de estudo maior tendem a apresentar um desempenho melhor. Ambas as visualizações exibem esse padrão com pequenas exceções. No grupo de estudantes com as 5% melhores médias gerais, há uma variação na sequência do grau de estudo. Os estudantes cuja mãe nunca estudou aparecem em maior quantidade sobre aqueles cuja mãe estudou até o quinto ano do Ensino Fundamental. A proporção de alunos nesses subgrupos é, no entanto, muito próxima. Outra exceção ocorre no grupo de estudantes com os piores desempenhos. Os estudantes cuja mãe possui pós-graduação completa aparecem em maior quantidade em relação aqueles cuja mãe possui superior completo.

A análise considerando o grau de estudo dos pais dos estudantes possui resultados semelhantes. Entre aqueles que obtiveram melhores resultados, a proporção dos subgrupos está diretamente ligada ao grau de estudo dos pais sem exceção. Estudantes cujo pai possui maior grau de estudo são mais frequentes que aqueles cujo pai possui um grau de estudo menor. No grupo que apresentou os piores resultados, desconsiderando os alunos que não sabem o grau de estudo do pai, a sequência proporcional foi a mesma que a apresentada na visualização referente ao grau de estudo



da mãe. A proporção também apresenta uma ligação com o nível do grau de estudo.

Quantidade de estudantes por grau de estudo da mãe

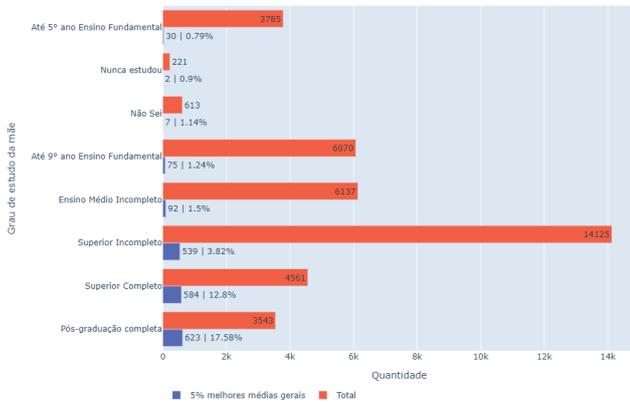


Figura 7. Melhores médias por grau de estudo da mãe

Quantidade de estudantes por grau de estudo da mãe

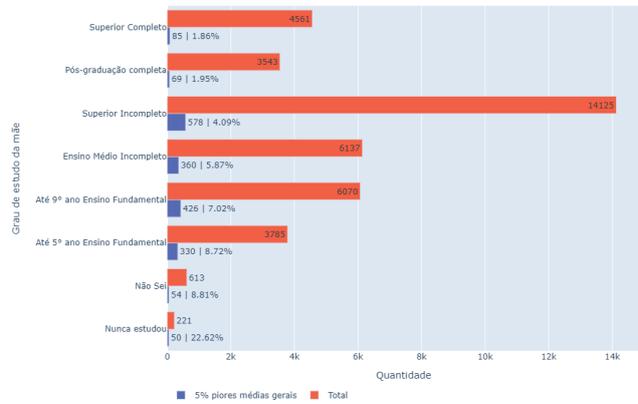


Figura 8. Piores médias por grau de estudo da mãe

O tipo de escola é analisado nas Figuras 9 e 10. As escolas privadas e federais recebem grande destaque no grupo de estudantes com os melhores desempenhos. Nas escolas privadas, quase um em cada cinco estudantes aparece entre as 5% melhores médias. Nas escolas federais, esse número é de um pouco mais de um em cada dez. As escolas estaduais representam quase 77% de todos os alunos que realizaram o exame, sendo essa a maior parcela entre os subgrupos. No entanto, esse subgrupo fica à frente apenas das escolas municipais entre os melhores resultados. Apenas 1,5% dos estudantes oriundos desse tipo de escola aparecem entre as melhores médias. Estudantes de escolas federais e privadas aparecem em menor fração em relação aos de escolas municipais e estaduais no grupo de estudantes com os piores resultados. Através dessa análise, pode-se avaliar que as escolas particulares e federais tiveram estudantes com melhor desempenho.

Quantidade de estudantes por tipo dependência administrativa escola

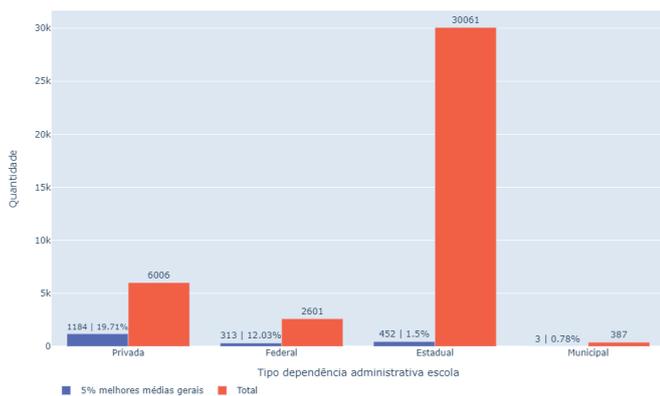


Figura 9. Melhores médias por tipo de escola

Quantidade de estudantes por tipo dependência administrativa escola

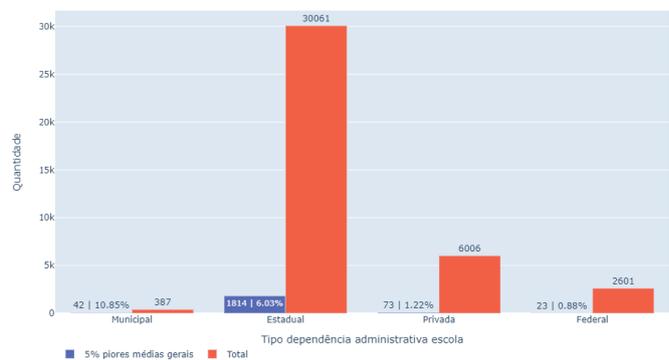


Figura 10. Piores médias por tipo de escola

Por fim, as Figuras 11 e 12 expõem a organização de estudantes, considerando o acesso à internet em casa. Possuir acesso a esse recurso não apresenta relação direta com o desempenho dos estudantes, haja vista que a proporção de estudantes desses subgrupos é semelhante entre os melhores e os piores resultados. Já a falta de acesso a esse recurso aparenta ter um impacto significativo no desempenho dos alunos. Menos de 1% dos



Quantidade de estudantes por acesso à Internet em casa

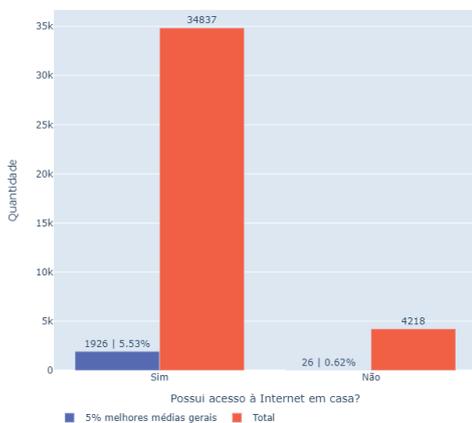


Figura 11. Melhores médias por acesso à internet

Quantidade de estudantes por acesso à Internet em casa

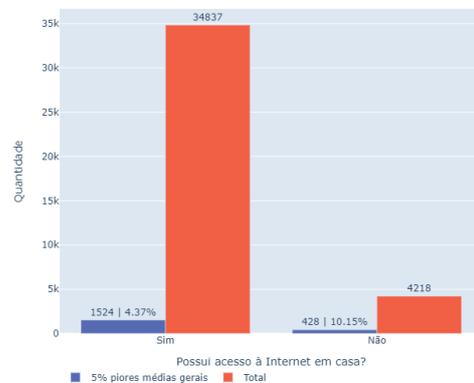


Figura 12. Piores médias por acesso à internet

estudantes enquadrados nesse contexto aparecem entre as melhores médias. Esse número é mais de cinco vezes menor que a proporção de estudantes que possuem acesso ao recurso em casa. Mais de 10% desses estudantes aparecem no grupo das 5% piores médias gerais. Essa proporção é maior que o dobro daqueles que possuem acesso ao recurso.

5. Conclusão

Este artigo analisou os dados dos estudantes do Rio Grande do Sul no ENEM 2019, visando identificar as características dos estudantes pertencentes aos grupos de 5% melhores e piores médias. A análise evidencia que o perfil socioeconômico influencia diretamente no desempenho dos estudantes, demonstrando a desigualdade educacional do Brasil. Estudantes de raça branca aparecem em maior proporção no grupo de 5% melhores médias gerais. Os resultados dos alunos tendem a aumentar conforme o aumento na renda familiar. Esse padrão ocorre devido à existência de uma tendência de menor investimento em educação em famílias com renda mais baixa. Essas famílias também possuem um acesso limitado à cultura e serviços, prejudicando o processo de aprendizagem dos estudantes. O acesso à internet, por exemplo, destacou-se como característica relacionada a um resultado pior. Menos de 1% dos estudantes sem acesso à internet aparecem entre os melhores desempenhos. A proporção de estudantes sem acesso à internet no grupo de piores médias ultrapassa 10%.

Além disso, um grau de estudo maior dos pais também é característica dos estudantes que se enquadram no grupo de melhores desempenhos. Em contrapartida, os alunos presentes no grupo de piores desempenhos possuem pais com grau menor de estudo ou, até mesmo, que nunca estudaram. A análise também realçou a diferença entre os tipos das escolas brasileiras. Alunos oriundos de escolas privadas e federais se destacam entre os melhores desempenhos, enquanto os oriundos de escolas municipais e estaduais aparecem em maior proporção no grupo de piores desempenhos. Essa é uma observação relevante, pois os alunos oriundos de escolas estaduais representam a maior parcela de participantes do exame. Porém, esses alunos não aparecem na mesma proporção no grupo de estudantes com melhores médias.

Como trabalhos futuros, os dados analisados podem ser comparados com dados de anos anteriores. Essa análise pode revelar mudanças de comportamento e perfil a



cada edição do exame. Os dados também podem ser comparados com outras regiões e estados, buscando detalhar ainda mais as diferenças socioeconômicas no Brasil. Outros atributos podem ser analisados, buscando identificar relações desses com o desempenho dos estudantes. Desta forma, os dados do ENEM se mostram uma fonte de dados valiosa para o estudo das características do ensino, das escolas e dos estudantes brasileiros.

Referências Bibliográficas

- R. Baker. *Data Mining for Education*. Elsevier, Oxford, UK, 3rd edition, 2014.
- A. P. Corti. As diversas faces do ENEM: análise do perfil dos participantes (1999-2007). *Estudos em Avaliação Educacional*, 24(55):198–221, 2013.
- M. F. C. de Paula. Políticas de democratização da educação superior brasileira: limites e desafios para a próxima década. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 22(2):301–315, 2017. ISSN 1414-4077. doi: 10.1590/s1414-40772017000200002.
- E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. V. Erven. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, 94(C):335–343, 2019. doi: 10.1016/j.jbusres.2018.02.
- J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. Elsevier, Waltham, 3 edition, 2012. ISBN 978-0-12-381479-1. doi: 10.1016/B978-0-12-381479-1.00001-0.
- Instituto Brasileiro de Geografia e Estatística. IBGE - Comitê de Estatísticas Sociais - Exame Nacional do Ensino Médio, 2020. URL <https://ces.ibge.gov.br/base-de-dados/metadados/inep/exame-nacional-do-ensino-medio-enem.html>. [Accessed 2020-10-30].
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Enem, 2020. URL <https://enem.inep.gov.br/>. [Accessed 2020-10-30].
- L. C. B. Martins, R. N. Carvalho, R. S. Carvalho, M. C. Victorino, and M. Holanda. Early prediction of college attrition using data mining. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1075–1078, 2017.
- S. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan. Mining in Educational Data: Review and Future Directions. In A. Hassanien, A. Azar, T. Gaber, D. Oliva, and F. Tolba, editors, *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, volume 1153, pages 92–102. Springer, Cham, 2020. ISBN 9783030442897. doi: 10.1007/978-3-030-44289-7.
- B. Santos, C. G. Oliveira, L. O. H. Topin, O. M. Mendizabal, and R. Barwaldt. Analysis of Candidates Profile for the National Entrance Exams for Admission to Brazilian Universities. *Proceedings - Frontiers in Education Conference, FIE*, 2019-October: 1–8, 2019. ISSN 15394565. doi: 10.1109/FIE43999.2019.9028381.
- E. Viggiano and C. Mattos. O desempenho de estudantes no ENEM 2010 em diferentes regiões brasileiras. *Revista Brasileira de Estudos Pedagógicos*, 94(237), 2013.
- C. G. Webber, G. Cini, and M. F. W. P. Lima. Facilitando a Análise de Dados Educacionais através de Ferramentas de Visualização. *Renote*, 11(3):1–10, 2014. ISSN 1679-1916. doi: 10.22456/1679-1916.44714.
- R. Zacchi, M. Ney, and N. Ponciano. Desigualdades Educacionais na Educação Básica: Uma Investigação a Partir do Exame Nacional do Ensino Médio. *Revista Vértices*, 18 (1):79–108, 2016. ISSN 1809-2667. doi: 10.19180/1809-2667.v18n116-05.