



## A Pandemia da COVID-19 impactou o ENEM? Uma Análise Comparativa de Dados dos Anos de 2019 e 2020

Nelson Weber Neto, UFMA, nelsonweberneto@gmail.com, 0000-0002-1136-768X

Raimundo C. Soares, UFMA, soares.raimundo@ufma.br, 0000-0002-0904-0370

Luciano Reis Coutinho, UFMA, luciano.rc@ufma.br, 0000-0001-7996-7334

Ariel Soares Teles, IFMA/UFMA, ariel.teles@ifma.edu.br, 0000-0002-0840-3870

**Resumo:** O Ministério da Educação do Brasil avalia o desempenho escolar dos estudantes ao término da educação básica através do Exame Nacional do Ensino Médio (ENEM), o qual é também usado para a entrada dos estudantes no ensino superior. Com a pandemia da COVID-19, muitas escolas tiveram que pausar as aulas ou manter o ensino de forma remota, o que afetou o processo de ensino e aprendizagem. Este artigo objetiva identificar se a pandemia impactou o ENEM, utilizando os microdados do ENEM de 2019, ano anterior ao início da pandemia, e do ENEM de 2020, primeiro ano da pandemia. O estudo realiza uma análise exploratória baseada em mineração de dados para comparar os dados dos dois anos. Os resultados apontam para uma melhoria nas notas médias gerais e aumento da disparidade de notas entre os candidatos de maior e menor renda.

**Palavras-chave:** Mineração de Dados, COVID-19, Pandemia, Educação Básica, Ensino Médio.

### Did the COVID-19 Pandemic impact ENEM? A Comparative Data Analysis for the Years 2019 and 2020

**Abstract:** The Brazilian Ministry of Education evaluates students' school performance at the end of basic education using the National High School Examination (NHSE), which is also used for students' entry into higher education. With the COVID-19 pandemic, many schools had to pause classes or maintain teaching remotely, which affected the teaching and learning process. This paper aims to identify whether the pandemic impacted NHSE, using microdata from the NHSE of 2019, the year before the beginning of the pandemic, and the NHSE of 2020, the first year of the pandemic. The study performs an exploratory data mining-based analysis to compare data from these two years. Results demonstrate an improvement in the general average grades and an increase in the disparity of grades between the candidates with higher and lower income.

**Keywords:** Data Mining, COVID-19, Pandemic, Basic Education, High School.

## 1. Introdução

A Organização Mundial da Saúde (OMS) declarou a COVID-19 como pandemia em 11 de março de 2020. O vírus da COVID-19 é altamente contagioso e causa sérios prejuízos aos seres humanos. Essas características tornaram o coronavírus uma das maiores pandemias da história, afetando todos os países. Com a chegada da pandemia da COVID-19, as instituições de ensino começaram a criar metodologias para que a educação não fosse prejudicada e o processo de ensino pudesse ocorrer. Entretanto, a pandemia da COVID-19 afetou diretamente a vida dos participantes de diferentes maneiras (Godoy *et al.*, 2021), independente do nível de ensino ou curso. O fechamento de escolas modificou a estrutura de aprendizagem e escolarização, bem como os métodos de ensino e avaliação (Todos Pela Educação, 2020).

Tendo em vista as mudanças causadas por essa pandemia, diversas soluções educacionais foram propostas visando preservar a continuidade do processo educacional. Algumas escolas conseguiram adotar aulas remotas e métodos de melhoria do



aprendizado através da utilização de recursos em ambientes digitais, método conhecido como *e-learning* e bastante explorado com o avanço dos ambientes de ensino online (Alqahtani e Rajkhan, 2020). Por outro lado, diversas escolas tiveram que seguir as recomendações de isolamento social, sem atividades presenciais por algum período, e sem conseguir manter o ensino dos alunos, mesmo de forma remota. O ensino remoto se tornou um desafio tanto para os estudantes quanto para os docentes (Gularte; Nascimento e Carvalho, 2021). Embora a educação remota tenha sido disponível antes da pandemia da COVID-19, esse tipo de ensino ainda era incomum nas instituições de ensino básico e a maioria das atividades de aprendizagem ocorria em salas de aula, presencialmente.

No Brasil, existem diversos programas de avaliação da educação, passando desde a avaliação do ensino infantil até o nível superior. Segundo o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), o Exame Nacional do Ensino Médio (ENEM) é uma avaliação criada e aplicada pelo Ministério da Educação (MEC) para medir o desempenho de alunos que concluíram o ensino médio (INEP, 2022a). Os microdados do ENEM reúnem um conjunto de informações detalhadas relacionadas às informações obtidos através desse exame. O ENEM é realizado todos os anos em diversos municípios do Brasil. Porém, com a chegada da pandemia, houve mudanças significativas na realização desse exame. O ENEM era normalmente realizado entre os meses de outubro e novembro, mas o exame referente ao ano de 2020 foi realizado em janeiro de 2021, devido à pandemia. Adicionalmente, outra mudança foi a chegada do ENEM Digital, um novo modelo opcional que permitiu o aluno fazer a mesma prova, mas em um ambiente virtual e gamificado.

Com essas mudanças ocasionadas pela pandemia da COVID-19, torna-se necessária a compreensão dos seus impactos na educação básica e como ela pode ter afetado no principal exame de desempenho do ensino médio no Brasil. Para identificar e analisar os impactos, técnicas de mineração de dados podem ser utilizadas. A mineração de dados é um processo computacional aplicado em diversas áreas, visando obter informações e gerar conhecimento (Peña-Ayala, 2014). A mineração de dados educacionais é um campo de pesquisa que explora a essência do conceito de mineração de dados para ajudar a descobrir informações úteis, tornando a utilização muito benéfica para o processo de pesquisa educacional (Peña-Ayala, 2014).

O objetivo desse estudo é utilizar os microdados do ENEM dos anos de 2019 e 2020 disponibilizados pelo INEP para comparar os resultados obtidos na prova através da utilização de mineração de dados. Este estudo visou responder a seguinte questão de pesquisa: *A pandemia da COVID-19 impactou o ENEM?* Para responder essa pergunta, foi feito o levantamento dos dados dos dois anos, pré-processamento deles e, por fim, uma análise exploratória e comparativa.

O restante deste artigo está organizado como segue. A Seção 2 discute os trabalhos relacionados. A Seção 3 descreve a metodologia utilizada neste estudo. A Seção 4 descreve e discute os resultados encontrados através do processo de mineração de dados realizado. Por fim, a Seção 5 relata as considerações finais.

## 2. Trabalhos Relacionados

As áreas de pesquisa relacionadas à ciência de dados evoluíram bastante, o que possibilitou o crescente número de trabalhos usando técnicas de mineração de dados para resolver problemas educacionais. Dentre os anos de 2006 a 2010, o número de trabalhos publicados sobre mineração de dados educacionais era de apenas quinze no Brasil (Rodrigues *et al.*, 2014). Já de 2011 até meados de 2014, foram publicados cerca de 50 artigos, mostrando uma evolução significativa. A revisão sistemática de (Soares *et*



al., 2021) foca nos trabalhos que utilizam técnicas de mineração de dados educacionais usando as bases disponibilizadas pelo INEP. Ela mostrou que, dos 19 trabalhos revisados, 11 deles focaram em problemáticas relacionadas ao desempenho acadêmico, englobando exames como o ENEM.

O trabalho de (Silva; Morino e Sato, 2014) utilizou os dados do ENEM do ano de 2010, usando uma tarefa da mineração de dados conhecida por Associação de Dados para encontrar padrões de regras para analisar relações de causa e efeito entre o desempenho dos alunos e fatores socioeconômicos das capitais da região Sudeste do Brasil. O trabalho de (Barcellos *et al.*, 2020) utiliza técnicas de mineração de dados nos microdados do ENEM de 2018, buscando entender a relação de aspectos financeiros como renda familiar e bens materiais na nota dos candidatos. Já o artigo de (Nakazone e Bortolotti, 2021) faz uma análise exploratória do ENEM dos anos de 2015 a 2019, para entender as diferenças entre as provas ao longo dos anos.

Por outro lado, diversos estudos ao redor do mundo têm focado nos impactos da pandemia da COVID-19 na educação. Por exemplo, para evidenciar a importância do tema, o editorial de (Reuge *et al.*, 2021) apresenta uma edição especial proposta pela *United Nations International Children's Emergency Fund* (UNICEF) em resposta à pandemia da COVID-19. O editorial apresenta brevemente 11 artigos publicados na edição especial, os quais estudam temas como o fechamento das escolas, a efetividade das soluções de ensino remoto, implicações de equidade, a mitigação de perda de aprendizagem, e noções sobre as melhores maneiras de reabrir as instituições de ensino.

O artigo de (Junior; Matos e Borges, 2021) propôs uma análise dos perfis dos estudantes durante a pandemia da COVID-19. Conforme os resultados do artigo, os alunos que possuíam perfil de melhores condições de acesso e infraestrutura, eram mais engajados nas atividades propostas com a metodologia de Educação à Distância (EaD). O estudo de (Moraes; Peres e Pedreira, 2021) buscou entender fatores e relações de variáveis apresentadas nos microdados do ENEM durante a pandemia, e avaliar o desempenho de alunos na área de conhecimento de matemática e suas tecnologias. O estudo mostrou que a diferença social existente entre os estudantes tende a permanecer, se os governantes e gestores não intervirem e criarem políticas públicas adequadas. Motivados pelos possíveis danos que a pandemia da COVID-19 poderia causar, (Sáiz-Manzanares *et al.*, 2021) propuseram a ferramenta *UBUMonitor*. Ela é uma aplicação que usa dados do *Moodle* para aplicar técnicas de mineração de dados educacionais para monitorar e detectar alunos em risco de evasão.

Conforme apresentado, já existem diversos trabalhos utilizando mineração de dados educacionais com o intuito de analisar e compreender problemas educacionais do Brasil. Entretanto, para o melhor do nosso conhecimento, não existe trabalho publicado na literatura científica que tenha conduzido um estudo de mineração de dados do ENEM de modo a comparar os anos de 2019 e 2020 para compreender o impacto da pandemia da COVID-19. Portanto, a contribuição deste estudo é a de realizar uma análise exploratória comparativa entre os dados do ano anterior à pandemia (2019) e o ano que iniciou a pandemia (2020) para identificar se houve impacto no ENEM.

### 3. Metodologia

A proposta de trabalho empregada neste estudo é a mineração de dados educacionais do ENEM de 2019 e 2020. Os microdados do ENEM ofertam dados educacionais de todos os estados e regiões do Brasil, por dependência administrativa (pública ou privada), ano de conclusão do ensino médio dos candidatos, e outras



informações pertinentes aos inscritos e às escolas que estudam ou estudaram. Para este trabalho, foram utilizados os dados de todo o Brasil.

Utilizamos a metodologia *Cross-Industry Standard Process for Data Mining* (CRISP-DM). A metodologia CRISP-DM descreve o ciclo de vida de um projeto de mineração de dados. Essa metodologia consiste em várias etapas iterativas, as quais são: entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e distribuição (Azevedo e Santos, 2008; Shearer, 2000). Neste trabalho, usamos a linguagem de programação *Python*, as bibliotecas *Pandas*, *NumPy*, e *Matplotlib*, com o *Google Colaboratory* como ambiente de desenvolvimento.

### 3.1. Entendimento do Negócio e dos Dados

A primeira etapa da metodologia CRISP-DM consiste em compreender o negócio e os dados, e então traduzir esse conhecimento na definição de um problema de mineração de dados. Utilizamos dois conjuntos de dados, os microdados do ENEM dos anos de 2019 e 2020, ambos disponibilizados pelo INEP em seu portfólio de dados abertos. Os dois conjuntos utilizados nesse estudo trazem dados sobre: o participante, a escola do participante, atendimento especializado para participantes que possuem alguma deficiência, o local de aplicação da prova, e dados sobre as provas objetiva, redação, e questionário socioeconômico.

O conjunto de dados do ENEM de 2019 possui 5.097.407 linhas e 136 colunas. Cada linha equivale a um candidato inscrito, mostrando informações do participante, como município onde a prova foi feita, escolaridade dos pais, dados referentes a alguma necessidade especial, dentre outros dados. Os microdados do ENEM de 2020, por sua vez, sofreram uma mudança. Os dados que antes eram formados por 136 colunas, agora passaram a ter apenas 76, contendo dados dos inscritos. O conjunto de dados possui 5.827.389 linhas, representando o número de participantes inscritos.

### 3.2. Preparação dos Dados

As tarefas dessa etapa contém a seleção de dados, limpeza, formatação e integração, gerando um novo conjunto de dados a ser utilizado como base para o estudo. Na etapa de preparação dos dados, foi criada mais uma coluna com a nota média dos participantes, pois o conjunto de dados só possuía apenas as médias de cada uma das áreas de conhecimento. Após a criação da coluna, foi feita a remoção das linhas que possuíam dados nulos na nota média, as quais representavam os candidatos faltantes. Houve também a transformação de colunas como raça, sexo e dependência administrativas para outras nomenclaturas, devido à utilização dessas características no processo de análise, tornando melhor a compreensão das informações.

Nesta etapa também foi utilizado o dicionário de microdados do ENEM, uma ferramenta disponibilizada pelo INEP para auxiliar no processo de entendimento das colunas a serem exploradas. Vale ressaltar que o processo de preparação de dados é essencial para a etapa de análise dos dados. Portanto, os conjuntos de dados dos anos de 2019 e 2020 passaram pelos mesmos processos de preparação.

### 3.3. Análise dos Dados

A análise de dados refere-se ao processo crítico de realizar investigações iniciais em dados para descobrir padrões, detectar anomalias, testar hipóteses e verificar suposições com a ajuda de estatísticas descritivas e representações gráficas (Komorowski *et al.*, 2016). O objetivo dessa etapa neste estudo foi extrair dados que pudessem ser efetivamente transformados em informações importantes para entender os impactos no ENEM causados pela pandemia. Além da análise exploratória dos dados, o principal foco



dessa etapa foi gerar informações comparativas e analisar se houve impacto (i.e., se houve diferenças).

Do conjunto total de dados, foram selecionados as colunas referentes às questões socioeconômicas, instituições de ensino, informações de realização da prova, e dados sobre o participante. Para diversas análises de informações, consideramos apenas os participantes presentes, o qual é caracterizado por estar presente nos dois dias de prova. Após a seleção das características abordadas, geramos análises exploratórias, com o uso de gráficos e tabelas, individualmente para cada um dos anos (i.e., dois *notebooks* no *Google Colaboratory*). Após a análise inicial de cada ano, geramos uma análise (i.e., um terceiro *notebook*), carregando os microdados dos dois anos para propósitos de comparação.

#### 4. Resultados e Discussão

A Figura 1 mostra a quantidade de inscrições no ENEM do ano de 2019 e 2020, exibindo a quantidade de ausentes (i.e., faltou um dia ou os dois) e presentes (i.e., realizou a prova os dois dias). A diferença de candidatos presentes entre os anos foi grande. No ano de 2019, 77% dos candidatos inscritos estiveram presentes, enquanto os presentes em 2020 foram cerca de 48% dos inscritos. Em números absolutos, o ENEM de 2019 teve 1.160.151 candidatos inscritos ausentes, enquanto o ENEM de 2020 teve 3.029.391 de ausentes. Houve também uma diferença no número de inscritos em 2020, com 5.827.389, cerca de 14% a mais. Vale ressaltar que no ano de 2020 também foi realizado o ENEM Digital. Porém, essa modalidade ainda implica na ida do candidato a uma instituição para a realização da prova.

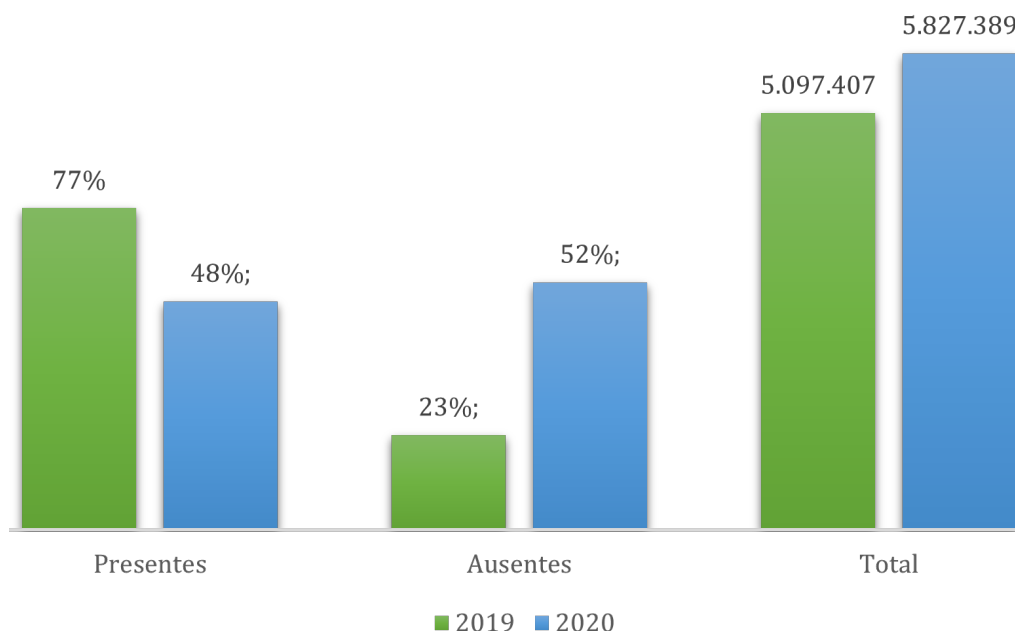


Figura 1. Inscritos presentes, ausentes e totais para as provas do ENEM dos anos de 2019 e 2020.

Outro dado interessante analisado foi a quantidade de reaplicações de provas solicitadas. Uma reaplicação é uma segunda prova do ENEM para os candidatos que tiveram problemas com a primeira aplicação. Elas são somente disponibilizadas para algumas situações, tais como: a ocorrência de incidentes logísticos (e.g., desastres naturais, falta de energia), falhas na entrega das provas aos locais de aplicação, e quando o participante possui alguma doença infectocontagiosa. No caso do ENEM de 2019, foram



solicitadas apenas 150 reaplicações, enquanto em 2020, com a chegada da COVID-19 como doença infectocontagiosa, houve um total de 235.204 solicitações. Vale ressaltar que, desses 235.204, estão inclusos também as aplicações do Estado do Amazonas, e os municípios de Rolim de Moura e Espigão D'Oeste, ambos em Rondônia, devido à justiça ter adiado a aplicação por conta de um surto de COVID-19.

A Figura 2 mostra a porcentagem dos inscritos do sexo masculino e feminino e suas respectivas raças declaradas. Pode-se notar que os números são próximos. A maior diferença está na quantidade de mulheres brancas que aumentou em 2020. Dentre as raças, houve pouca diferença entre os anos, e a diferença entre participantes pardos ou brancos, em relação às demais raças, continuou bastante significativa. No que tange ao sexo, não houve mudança no número entre homens e mulheres de um ano para outro: a porcentagem de participantes do sexo masculino ficou em cerca de 40%, enquanto a porcentagem das participantes ficou próximo de 60%.

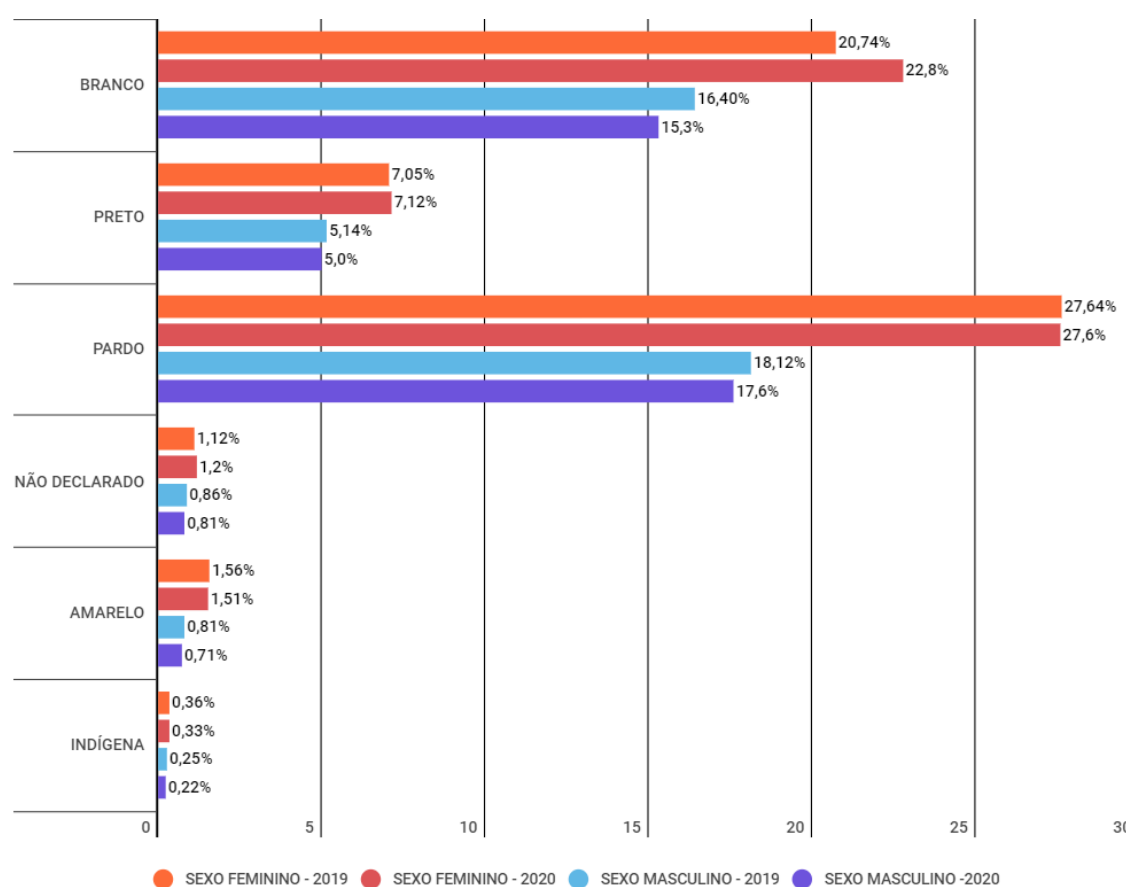


Figura 2. Quantidade de participantes considerando cor, raça e sexo no ENEM dos anos de 2019 e 2020.

A Figura 3 apresenta um mapa de calor com a diferença entre as notas médias dos anos: *notas médias de 2020 – notas médias de 2019*. Portanto, para esse gráfico, foi feita a subtração das notas médias de cada área do conhecimento do ano de 2020 pelas notas médias das suas respectivas áreas do ano de 2019. Pode-se notar que os estados tiveram um aumento sutil nas diversas áreas do conhecimento, exceto na área da matemática e suas tecnologias, que teve diminuição em 21 estados, e redação, que diminuiu em 14 estados. A área de conhecimento que teve somente aumento de notas médias foi a ciências da natureza. De todos os estados do país, 22 tiveram um aumento leve em suas notas médias. Os estados de São Paulo, Minas Gerais e Paraná tiveram os maiores aumentos na nota média. Apesar dos aumentos, não é possível concluir que





houve uma melhora na educação básica brasileira, uma vez que o aumento na nota média nacional foi de apenas 2,55 pontos.



Figura 3. Diferença das Médias das áreas de conhecimento do ENEM dos anos de 2019 e 2020 em relação aos Estados do Brasil. Legendas: LC - Linguagens e Códigos; MTM - Matemática; RED - Redação; CN - Ciências da Natureza; CH - Ciências Humanas; M - Média.

A Figura 4 apresenta um *box plot* das notas dos participantes considerando a categoria de dependência administrativa da instituição que estudam ou, em caso de participantes concluintes, que estudaram. Observa-se que as medianas das escolas privadas são maiores em ambos os anos. Ao comparar as categorias de dependência, essa ordem de classificação é seguida pelas escolas federais, municipais e estaduais. O gráfico ainda mostra que, das escolas públicas, as escolas de dependência federal têm mediana maior, sendo muito próxima ao ensino privado. Em relação aos *outliers*, pode-se notar que as escolas estaduais, seguida das privadas e depois federais, possuem os maiores números, acima dos limites máximos e mínimos. Isso mostra que existem várias instituições que se destacam (positivamente, mas também negativamente) em relação ao centro de distribuição dos dados. As escolas de dependência municipal são as que possuem menor quantidade de *outliers*.

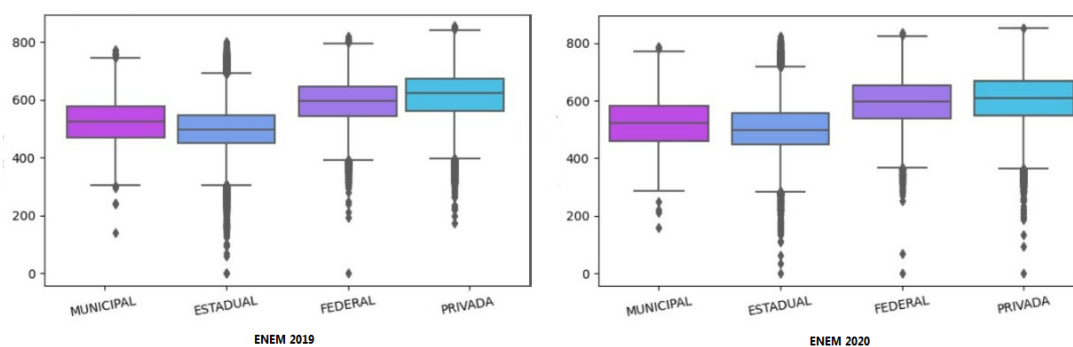


Figura 4. Gráfico de *box plot* dos participantes por categoria de dependência administrativa da instituição de ensino.

A disparidade entre as notas dos candidatos que vêm de escolas públicas e privadas é mais visível ao analisar a ordem de classificação das instituições. No ano de 2019, as sete primeiras escolas com maior nota média são privadas. Nessa lista, só a oitava posição é ocupada por uma escola federal. Já as escolas estaduais e municipais só aparecem na posição de número 166 e 607, respectivamente. Em 2020 não foi possível realizar essa análise, dado que o INEP removeu o código da escola do conjunto de dados, considerando que há evidências de que disponibilizar a informação do código da escola permite, em alguns casos, a identificação dos participantes por parte dos gestores dessas instituições” (INEP, 2022b).

Para analisar as características que tiveram maior influência na nota média do ENEM, utilizamos a correlação de Pearson (Benesty *et al.*, 2009). A Tabela 1 mostra os



resultados da correlação, onde pode-se observar que, nos dois anos, a renda familiar é a variável com maior correlação com a nota média dos participantes, seguida da quantidade de computadores na residência. Ainda pode-se notar que entre os anos de 2019 e 2020, a variável que sofreu maior alteração foi relacionada a possuir Internet na residência (“Acesso à Internet” na Tabela 1): a correlação em 2019 era 0,26, passou a ser de 0,40 em 2020, tornando-se uma variável de correlação moderada.

Tabela 1. Maiores correlações entre as variáveis do questionário socioeconômico e a nota média dos candidatos.

| Questão Socioeconômica | ENEM2019 | ENEM2020 |
|------------------------|----------|----------|
| Renda Familiar         | 0,48     | 0,51     |
| Possui Computadores    | 0,41     | 0,46     |
| Possui Celulares       | 0,28     | 0,32     |
| Acesso à Internet      | 0,26     | 0,40     |
| Escolaridade Pai       | 0,25     | 0,29     |
| Escolaridade Mãe       | 0,31     | 0,34     |

Dado que a renda familiar do candidato foi a variável de maior correlação com nota média, criamos a Tabela 2, a qual apresenta a nota média geral dos candidatos de cada classe dessa variável. Ao analisar os dados, pode-se observar uma disparidade das notas médias entre as diferentes classes de renda familiar. No ano de 2019, houve uma diferença entre a nota média dos candidatos que responderam “Nenhuma Renda” - 468,70 - e aqueles participantes com maior renda (acima de R\$ 19.960,00 em 2019) - 644,88 - que foi de 176,18 pontos. No ano de 2020, essa diferença aumentou para 190,2 pontos. Observa-se também que, no ano de 2020, a nota média dos candidatos aumentou levemente em todas as classes de renda, com exceção da classe “Nenhuma Renda”, a qual teve uma redução de 6,8 pontos. Além disso, esse aumento de notas foi menor nas classes de menor renda: as 3 classes imediatamente acima da classe “Nenhuma Renda” tiveram aumento de apenas 0,46, 4,54 e 0,68, respectivamente. Isso mostra um impacto da pandemia da COVID-19, que prejudicou os candidatos de menor renda.

Tabela 2. Renda Familiar e Nota Média dos Participantes.

| Renda Familiar                  | ENEM 2019 | Renda Familiar                  | ENEM 2020 |
|---------------------------------|-----------|---------------------------------|-----------|
| Nenhuma Renda                   | 468,70    | Nenhuma Renda                   | 461,90    |
| Até R\$ 998,00                  | 479,24    | Até R\$1,045,00                 | 479,70    |
| R\$ 998,01 até R\$ 1.497,00     | 499,72    | R\$ 1.045,01 até R\$ 1.567,50   | 504,26    |
| R\$ 1.497,01 até R\$ 1.996,00   | 518,76    | R\$ 1.567,51 até R\$ 2.090,00   | 519,44    |
| R\$ 1.996,01 até R\$ 2.495,00   | 526,80    | R\$ 2.090,01 até R\$ 2.612,50   | 537,38    |
| R\$ 2.495,01 até R\$ 2.994,00   | 544,04    | R\$ 2.612,51 até R\$ 3.135,00   | 549,28    |
| R\$ 2.994,01 até R\$ 3.992,00   | 553,30    | R\$ 3.135,01 até R\$ 4.180,00   | 568,04    |
| R\$ 3.992,01 até R\$ 4.990,00   | 571,80    | R\$ 4.180,01 até R\$ 5.225,00   | 584,68    |
| R\$ 4.990,01 até R\$ 5.988,00   | 581,56    | R\$ 5.225,01 até R\$ 6.270,00   | 596,92    |
| R\$ 5.988,01 até R\$ 6.986,00   | 592,96    | R\$ 6.270,01 até R\$ 7.315,00   | 606,58    |
| R\$ 6.986,01 até R\$ 7.984,00   | 598,90    | R\$ 7.315,01 até R\$ 8.360,00   | 612,94    |
| R\$ 7.984,01 até R\$ 8.982,00   | 607,48    | R\$ 8.360,01 até R\$ 9.405,00   | 617,17    |
| R\$ 8.982,01 até R\$ 9.980,00   | 610,77    | R\$ 9.405,01 até R\$ 10.450,00  | 622,86    |
| R\$ 9.980,01 até R\$ 11.976,00  | 618,38    | R\$ 10.450,01 até R\$ 12.540,00 | 629,26    |
| R\$ 11.976,01 até R\$ 14.970,00 | 627,89    | R\$ 12.540,01 até R\$ 15.675,00 | 636,72    |
| R\$ 14.970,01 até R\$ 19.960,00 | 635,80    | R\$ 15.675,01 até R\$ 20.900,00 | 644,66    |
| Acima de R\$ 19.960,00          | 644,88    | Acima de R\$ 20.900,00          | 652,10    |

## 5. Considerações Finais

Este estudo, através da utilização de mineração de dados educacionais, identificou informações relevantes para responder se a pandemia da COVID-19 impactou no ENEM. Para isso, uma análise exploratória e comparativa foi realizada com os microdados do





ENEM dos anos de 2019 e 2020. Identificamos que a presença de inscritos no ENEM de 2020 sofreu um grande impacto devido à pandemia. Isso ocorreu, possivelmente, devido às mudanças provocadas pela pandemia, tais como a mudança na data da prova, falta de preparação adequada dos candidatos, receio dos inscritos em contrair o coronavírus, e questões econômicas dos estudantes. Em relação às diferenças na raça dos participantes, notou-se que não existiu impacto ocasionado pela pandemia. No que diz respeito ao desempenho das instituições de ensino, as notas das escolas públicas, independente da sua dependência administrativa (municipal, estadual ou federal), e privadas não sofreram grandes alterações. Outra informação relevante encontrada está relacionada à diferença das notas médias dos estados que, apesar delas terem sofrido pequenas oscilações positivas ou negativas, as diferenças são muito pequenas. Os estados que em 2019 ocuparam as primeiras posições no ranqueamento do ENEM, permaneceram em 2020, o que mostra que, mesmo com o ensino remoto, ainda conseguiram manter suas posições. Por fim, um dado importante foi que a nota média de 2020 teve um aumento na correlação positiva com a variável relacionada ao candidato possuir acesso à Internet em casa. Isso demonstra que, devido ao ensino ter passado a ser remoto para diversas instituições de ensino, essa variável passou a influenciar mais nas notas dos candidatos.

A pesquisa desenvolvida e apresentada neste artigo ainda é inicial, por considerar apenas o primeiro ano de pandemia e, conseqüentemente, ela possui limitações. Como a pandemia da COVID-19 foi além do ano de 2020, e como seus prejuízos podem ter reflexos a longo prazo, trabalhos futuros deverão focar na investigação acerca dos impactos nos anos seguintes para comparar com os dados dos anos apresentados neste estudo, passando então a se tornar uma análise de série temporal. Além disso, os dados aqui apresentados levantaram uma série de outras questões de pesquisa que devem ser compreendidas através de uma análise mais profunda e individualizada (e.g., por estado, instituição, dependência administrativa). Essa investigação profunda dos dados trará conhecimentos científicos de suma importância, visto que os dados geográficos, demográficos e socioeconômicos, quando analisados individualizadamente, permitirão a obtenção de informações que podem ser utilizadas para o auxílio no processo de tomada de decisões baseado em evidências e criação de políticas públicas.

### Agradecimentos

Os autores gostariam de agradecer a Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) pelo apoio dado a seus projetos de pesquisa.

### Referências

- Alqahtani, A. Y.; Rajkhan, A. A. E-learning critical success factors during the COVID-19 pandemic: A comprehensive analysis of e-learning managerial perspectives. **Education Sciences**, MDPI, v. 10, n. 9, p. 216, ago. 2020.
- Azevedo, A. I. R. L.; Santos, M. F. Kdd, semma and crisp-dm: a parallel overview. **IADS-DM**, 2008.
- Barcellos, A. A.; Isotani, S.; Diego, C.; Damasceno, N. Mineração de dados abertos - enem 2018. **Anais dos Trabalhos de Conclusão de Curso da Pós-Graduação em Computação Aplicada à Educação**, Universidade São Paulo, 2020.
- Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In: **Noise Reduction in Speech Processing**. [S.l.]: Springer Berlin Heidelberg, 2009. p. 1–4.
- Godoy, L. D.; Falcoski, R.; Incrocci, R. M.; Versuti, F. M.; Padovan-Neto, F. E. The psychological impact of the COVID-19 pandemic in remote learning in higher education.



**Education Sciences**, MDPI, v. 11, n. 9, p. 473, 2021.

Gularte, F. N.; Nascimento, F. K. V. do; Carvalho, A. V. de. ENEM em tempos de pandemia: desafios enfrentados pelos docentes frente o ensino remoto da rede estadual de Presidente Kennedy/TO. **Research, Society and Development**, v. 10, n. 14, p. e5711101422339, nov. 2021.

INEP. **Exame Nacional do Ensino Médio (ENEM)**. 2022. <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>>. Online; Acessado em 06 de maio de 2022.

INEP. **Nota de esclarecimento — Divulgação dos microdados**. 2022. <<https://www.gov.br/inep/pt-br/assuntos/noticias/institucional/nota-de-esclarecimento-divulgacao-dos-microdados>>. Online; Acessado em 07 de maio de 2022.

Junior, L. P.; Matos, S. N.; Borges, H. B. Análise dos perfis de alunos do ensino superior sobre a realização de aulas na modalidade a distância durante pandemia da covid-19 usando algoritmos de aprendizagem de máquina. **RENOTE**, UFRGS, v. 18, n. 2, p. 336–345, jan. 2021.

Komorowski, M.; Marshall, D. C.; Salciccioli, J. D.; Crutain, Y. Exploratory data analysis. **Secondary Analysis of Electronic Health Records**, Springer International Publishing, p. 185–203, 2016.

Moraes, C. P. de; Peres, R. T.; Pedreira, C. E. Eficácia escolar e variáveis familiares em tempos de pandemia: um estudo a partir de dados do enem. **Interfaces da educação**, v. 12, n. 35, p. 635–658, nov. 2021.

Nakazone, E.; Bortolotti, L. M. Análise de dados históricos do enem entre 2015 à 2019. In: **Congresso de Tecnologia-Fatec Mococa**. [S.l.: s.n.], 2021. v. 4, n. 1.

Peña-Ayala, A. Educational data mining: A survey and a data mining-based analysis of recent works. **Expert Systems with Applications**, v. 41, n. 4, Part 1, p. 1432–1462, 2014.

Reuge, N. *et al.* Education response to COVID 19 pandemic, a special issue proposed by UNICEF: Editorial review. **International Journal of Educational Development**, Elsevier BV, v. 87, p. 102485, nov. 2021.

Rodrigues, R. L.; Ramos, J. L. C.; Silva, J. C. S.; Gomes, A. S. A literatura brasileira sobre mineração de dados educacionais. **Workshops do Congresso Brasileiro de Informática na Educação**, Sociedade Brasileira de Computação - SBC, nov. 2014.

Sáiz-Manzanares, M. C. *et al.* Monitoring of student learning in learning management systems: An application of educational data mining techniques. **Applied Sciences**, MDPI, v. 11, n. 6, p. 2677, 2021.

Shearer, C. The crisp-dm model: the new blueprint for data mining. **Journal of data warehousing**, THE DATA WAREHOUSE INSTITUTE, v. 5, n. 4, p. 13–22, 2000.

Silva, L. A.; Morino, A. H.; Sato, T. M. C. Prática de mineração de dados no exame nacional do ensino médio. In: **Workshops do Congresso Brasileiro de Informática na Educação**. [S.l.]: Sociedade Brasileira de Computação - SBC, 2014.

Soares, R. D. C. *et al.* Mineração de dados da educação básica brasileira usando as bases do INEP: Uma revisão sistemática da literatura. **RENOTE**, UFRGS, v. 19, n. 1, p. 361–370, jul. 2021.

Todos Pela Educação. **Ensino a distância na educação básica frente à pandemia da Covid-19**. 2020. <[https://www.todospelaeducacao.org.br/\\_uploads/\\_posts/425.pdf](https://www.todospelaeducacao.org.br/_uploads/_posts/425.pdf)>. Online; Acessado em 05 de maio de 2022.