



# **UNIVERSIDADE FEDERAL DO MARANHÃO**

**Programa de Pós-Graduação em Ciência da  
Computação**

**Nelson Weber Neto**

***Minerando Dados para Entender o Impacto da  
Pandemia da COVID-19 no Exame Nacional do  
Ensino Médio***

**São Luís**

**2023**

MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO MARANHÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**Minerando Dados para Entender o Impacto da  
Pandemia da COVID-19 no Exame Nacional do  
Ensino Médio**

**Nelson Weber Neto**

**São Luís, 2023**

Nelson Weber Neto

# **Minerando Dados para Entender o Impacto da Pandemia da COVID-19 no Exame Nacional do Ensino Médio**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação - da Universidade Federal do Maranhão, como requisito para obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Ariel Soares Teles

UFMA/IFMA

Coorientador: Prof. Dr. Luciano Reis Coutinho

UFMA

São Luís

2023

---

Nelson Weber Neto

Minerando Dados para Entender o Impacto da Pandemia da COVID-19 no Exame Nacional do Ensino Médio/ Nelson Weber Neto. – São Luís, 2023.

79 f.

Orientador: Prof. Dr. Ariel Soares Teles

Dissertação (Mestrado) – Universidade Federal do Maranhão – UFMA  
Programa de Pós-Graduação em Ciência da Computação, 2023.

1. Mineração de dados. 2. Educação. 3. ENEM. I. Soares Teles, Ariel, orient.  
II. Título.

CDU XXX

---

Nelson Weber Neto

## **Minerando Dados para Entender o Impacto da Pandemia da COVID-19 no Exame Nacional do Ensino Médio**

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Nelson Weber Neto e aprovada pela comissão examinadora.

Aprovada em 28 de abril de 2023.

### BANCA EXAMINADORA

---

Prof. Dr. Ariel Soares Teles (Orientador)  
UFMA/IFMA

---

Prof. Dr. Luciano Reis Coutinho (Coorientador)  
UFMA

---

Prof. Dr. Luis J. Enrique Rivero Cabrejos  
(Examinador Interno)  
UFMA

---

Prof. Dra. Anarosa Alves Franco Brandão  
(Examinadora Externa) USP

*Gostaria de dedicar a Deus, a toda a minha família, minha noiva e amigos que sempre acreditaram em mim, me apoiarem e não deixaram eu desistir nos momentos difíceis.*

# Agradecimentos

Agradeço imensamente a Deus por ter me abençoado e ter me dado saúde para este momento;

Agradeço também a minha mãe, meu pai e minha irmã por todo apoio dado durante esse período;

Agradeço a minha noiva por me apoiar nos momentos difíceis, por sempre estar ao meu lado nessa jornada e não deixar eu desistir nos momentos de aflição;

Ao meu orientador e co-orientador por toda paciência, ensinamentos e direcionamentos.

*"O conhecimento serve para encantar as pessoas, não para humilhá-las."  
(Mario Sergio Cortella)*



# Resumo

No Brasil, o principal exame de avaliação do desempenho da educação básica é o Exame Nacional do Ensino Médio (ENEM), que também é utilizado para o ingresso de alunos no ensino superior. Em 2020, com a chegada da COVID-19, as instituições de educação básica precisaram mudar seu modelo educacional com ensino presencial para a utilização de metodologias de ensino à distância, o que pode ter afetado a qualidade da educação recebida. Portanto, há uma necessidade por entender os efeitos que a pandemia da COVID-19 causaram ao ENEM. Esta dissertação de mestrado tem o objetivo de identificar os principais impactos causados pela pandemia no ENEM, considerando todo o Brasil. Para tanto, foi realizado o processo completo de Mineração de Dados Educacionais (MDE), baseado na metodologia *Cross Industry Standard Process for Data Mining* (CRISP-DM), no ENEM em 5 anos diferentes, a fim de compreender os anos pré-pandemia e os dois primeiros anos da pandemia. Em particular, esse estudo usou técnicas de análise exploratória descritiva, correlação e agrupamento hierárquico para identificar os impactos causados no ENEM. Os resultados mostram que os participantes com maiores rendas tiveram melhor desempenho, as escolas privadas tiveram um desempenho superior aos outros tipos de escolas, o número de presentes e ausentes no exame foi menor nos anos da pandemia, e houve uma mudança nas características socioeconômicas dos participantes. Por fim, a pandemia não impactou negativamente no desempenho dos estudantes, mas as características dos participantes que realizaram o exame mudaram e o número de ausentes e presentes no exame foi drasticamente impactado.

**Palavras-chave:** Mineração de Dados Educacionais, COVID-19, Educação, ENEM, Pandemia, Análise de Dados.

# Abstract

In Brazil, the main performance evaluation exam in basic education is the National Secondary Education Examination (ENEM), which is also used to enroll students in higher education. In 2020, with the arrival of COVID-19, basic education institutions needed to change their educational model from face-to-face teaching to the use of distance learning methodologies, which may have affected the quality of education received. Therefore, there is a need to understand the effects that the COVID-19 pandemic caused to ENEM. This master's thesis aims to identify the main impacts caused by the pandemic on ENEM, considering all of Brazil. To this end, the complete process of Educational Data Mining (MDE) was carried out, based on the *Cross Industry Standard Process for Data Mining* (CRISP-DM) methodology, in ENEM in 5 different years, in order to understand the years pre-pandemic and the first two years of the pandemic. In particular, this study used techniques of descriptive exploratory analysis, correlation and hierarchical grouping to identify the impacts caused on the ENEM. The results show that participants with higher incomes performed better, private schools outperformed other types of schools, the number of present and absentees in the exam was lower in the pandemic years, and there was a change in the socioeconomic characteristics of the participants. . Finally, the pandemic did not negatively impact student performance, but the characteristics of the participants who took the exam changed and the number of absentees and present in the exam was drastically impacted.

**Keywords:** Data mining, COVID-19, Education, ENEM, Pandemic, Data analysis.

# Lista de ilustrações

Figura 1 – Quantidade de casos de COVID-19 por região durante a décima terceira e a décima nona semana de 2020. . . . .	21
Figura 2 – Etapas do ciclo de vida dos dados . . . . .	27
Figura 3 – Tarefas de Mineração de Dados Educacionais (MDE) . . . . .	28
Figura 4 – Ciclo da vida do Cross Industry Standard Process for Data Mining (CRISP-DM) . . . . .	35
Figura 5 – Inscritos presentes, ausentes e totais para as provas do Exame Nacional do Ensino Médio (ENEM) nos 5 anos analisados. . . . .	44
Figura 6 – Desempenho dos participantes conforme o tipo de dependência administrativa durante os três anos de pré-pandemia analisados. . . . .	45
Figura 7 – Desempenho dos participantes conforme o tipo de dependência administrativa durante a pandemia da COVID-19. . . . .	46
Figura 8 – Notas dos participantes do ENEM conforme a renda familiar durante o período de pré-pandemia. . . . .	47
Figura 9 – Notas dos participantes do ENEM conforme a renda familiar durante o período da pandemia. . . . .	48
Figura 10 – Notas da área de ciências naturais e suas tecnologias dos cinco anos analisados. . . . .	49
Figura 11 – Notas da área de ciências humanas e suas tecnologias dos cinco anos analisados. . . . .	49
Figura 12 – Notas área de Linguagens, códigos e suas tecnologias dos cinco anos analisados. . . . .	50
Figura 13 – Notas da área de Matemática e suas tecnologias dos cinco anos analisados. . . . .	51
Figura 14 – Notas da Redação dos cinco anos analisados. . . . .	52
Figura 15 – Correlação das variáveis socioeconômicas com as notas médias dos cinco anos analisados. . . . .	52
Figura 16 – Mapas do Brasil com a nota média de cada estado nos 5 anos analisados. . . . .	54
Figura 17 – Série temporal do desempenho médio dos estados do Ceará, Maranhão e Piauí durante os 5 anos analisados. . . . .	55
Figura 18 – Dendrogramas dos agrupamentos hierárquicos dos cinco anos analisados. . . . .	58
Figura 19 – Frequência dos participantes de cada grupo em relação à nota média do ENEM no ano de 2017. . . . .	60
Figura 20 – Frequência dos participantes de cada grupo em relação à nota média do ENEM no ano de 2018. . . . .	61
Figura 21 – Frequência dos participantes de cada grupo em relação à nota média do ENEM no ano de 2019. . . . .	62

Figura 22 – Frequência dos participantes de cada grupo em relação à nota média do ENEM no ano de 2020. . . . .	63
Figura 23 – Frequência dos participantes de cada grupo em relação à nota média do ENEM no ano de 2021. . . . .	64

# Lista de tabelas

Tabela 1 – Histórico do ENEM. . . . .	24
Tabela 2 – Principais mudanças realizadas na prova em 2009 (TRAVITZKI, 2013). . . . .	24
Tabela 3 – Eixos Cognitivos do ENEM (CAMPOS, 2015). . . . .	25
Tabela 4 – Descrição das variáveis utilizadas. . . . .	38
Tabela 5 – Desempenho geral dos participantes por região no ENEM de 2017. . . . .	55
Tabela 6 – Desempenho geral dos participantes por região no ENEM de 2018. . . . .	56
Tabela 7 – Desempenho geral dos participantes por região no ENEM de 2019. . . . .	56
Tabela 8 – Desempenho geral dos participantes por região no ENEM de 2020. . . . .	57
Tabela 9 – Desempenho geral dos participantes por região no ENEM de 2021. . . . .	57
Tabela 10 – Agrupamento hierárquico do ENEM de 2017. . . . .	59
Tabela 11 – Agrupamento hierárquico do ENEM de 2018. . . . .	60
Tabela 12 – Agrupamento hierárquico do ENEM de 2019. . . . .	61
Tabela 13 – Agrupamento hierárquico do ENEM de 2020. . . . .	62
Tabela 14 – Agrupamento hierárquico do ENEM de 2021. . . . .	64

# Lista de Siglas

**CNE** Conselho Nacional de Educação.

**CRISP-DM** Cross Industry Standard Process for Data Mining.

**EAD** Educação a Distância.

**ENADE** Exame Nacional de Desempenho de Estudantes da Educação Superior.

**ENEM** Exame Nacional do Ensino Médio.

**FIES** Fundo de Financiamento ao Estudante do Ensino Superior.

**GPU** Graphics Processing Unit.

**IDEB** Índice de Desenvolvimento da Educação Básica.

**INEP** Instituto Nacional de Estudos Pedagógicos Anísio Teixeira.

**LGPD** Lei Geral de Proteção de Dados.

**MDE** Mineração de Dados Educacionais.

**MEC** Ministério da Educação.

**OCDE** Organização para a Cooperação e Desenvolvimento.

**OMS** Organização Mundial de Saúde.

**PISA** Programa Internacional de Avaliação de Alunos.

**PLN** Processamento de Linguagem Natural.

**PPGCC** Programa de Pós-graduação em Ciência da Computação.

**PROUNI** Programa Universidade Para Todos.

**REUNI** Reestruturação e Expansão das Universidades Federais.

**RSL** Revisão Sistemática de Literatura.

**SAEB** Sistema de Avaliação da Educação Básica.

**SISU** Sistema de Seleção Unificada.

**TPU** Tensor Processing Unit.

**TRI** Teoria de resposta ao item.

**UNESCO** Organização das Nações Unidas para a Educação, a Ciência e a Cultura.

**UNICEF** United Nations Children's Fund.

# Sumário

	Lista de tabelas . . . . .	xi
	Lista de Siglas . . . . .	xii
1	<b>INTRODUÇÃO</b> . . . . .	15
1.1	Contexto Geral . . . . .	15
1.2	Caracterização do Problema . . . . .	16
1.3	Hipótese de Pesquisa . . . . .	17
1.4	Relevância do Trabalho . . . . .	17
1.5	Objetivos . . . . .	18
1.6	Organização do Trabalho . . . . .	18
2	<b>FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	20
2.1	COVID-19 e a Educação Básica Brasileira . . . . .	20
2.2	ENEM . . . . .	23
2.3	Mineração de Dados Educacionais . . . . .	26
2.4	Conclusão . . . . .	29
3	<b>TRABALHOS RELACIONADOS</b> . . . . .	30
3.1	Mineração de Dados do ENEM . . . . .	30
3.2	Mineração de Dados Educacionais e COVID-19 . . . . .	32
3.3	Discussão . . . . .	33
4	<b>METODOLOGIA</b> . . . . .	34
4.1	CRISP-DM . . . . .	34
4.2	Questões de Pesquisa e Critérios de Sucesso . . . . .	35
4.3	Entendimento do Negócio e dos Dados do ENEM . . . . .	37
4.4	Ferramentas Utilizadas . . . . .	38
4.5	Preparação dos Dados . . . . .	40
4.6	Visualização e Modelagem dos Dados . . . . .	41
4.7	Conclusão . . . . .	42
5	<b>RESULTADOS</b> . . . . .	43
5.1	Impacto no Número de Participantes . . . . .	43
5.2	Impacto no Desempenho Considerando a Dependência Administra- tiva da Escola . . . . .	44
5.3	Impacto no Desempenho Considerando a Renda Familiar . . . . .	46



5.4	Áreas do Conhecimento Mais Impactadas . . . . .	48
5.5	Correlação das Características Socioeconômicas com o Desempenho	51
5.6	Impacto no Desempenho das Regiões e Estados . . . . .	53
5.7	Impactos nas Características dos Participantes . . . . .	57
6	<b>DISCUSSÃO</b> . . . . .	65
6.1	Principais Achados . . . . .	65
6.2	Verificação dos Critérios de Sucesso . . . . .	67
6.3	Limitações . . . . .	68
6.4	Trabalhos Futuros . . . . .	68
7	<b>CONSIDERAÇÕES FINAIS</b> . . . . .	70
7.1	Contribuições Científicas . . . . .	70
7.2	Publicações . . . . .	71
	<b>REFERÊNCIAS</b> . . . . .	73

# 1 Introdução

## 1.1 Contexto Geral

No final de 2019, uma pneumonia aguda com graves dificuldades apareceu em Wuhan, uma das maiores cidades da China. A doença foi reconhecida pela tecnologia de sequenciamento do genoma como uma síndrome respiratória aguda grave, sendo denominada de coronavírus ou COVID-19 (KHAN *et al.*, 2021). Com comportamento errático, ou seja, sem previsibilidade e com poder de rápida transmissão pelo mundo, a COVID-19 tornou-se uma emergência de saúde pública. No dia 11 do mês de março de 2020, a [Organização Mundial de Saúde \(OMS\)](#) declarou a COVID-19 como uma pandemia. Suas altas taxas de contágio e sua capacidade de causar sérios problemas de saúde, como falta de ar, fibrose nos pulmões, perda do paladar e do olfato, agravamento de doenças preexistentes, dentre outros problemas, fizeram com que a COVID-19 se tornasse uma das maiores pandemias da história, afetando todos os países (SURYASA; RODRÍGUEZ-GÁMEZ; KOLDORIS, 2021). Segundo (ROBBINS, 2020), em 4 de março, a [OMS](#) confirmou que 22 países anunciaram ou iniciaram o fechamento de escolas para impedir a propagação do novo coronavírus.

No Brasil, o vírus também se espalhou rapidamente e permeou em todos os estados brasileiros desde o surgimento do primeiro caso. Até o dia 31 de janeiro de 2023, o número estimado de pessoas com a doença chegou a 36.824.580, e 697.074 de mortos segundo (Ministério da Saúde, 2023). Obviamente, como em muitos outros aspectos da vida cotidiana, a COVID-19 teve um impacto severo em alunos, professores e instituições educacionais. As instituições de ensino criaram metodologias para que a educação não fosse comprometida e o processo de ensino pudesse ser realizado (EDUCAÇÃO, 2020). Vale ressaltar que esse não foi o primeiro fechamento de escolas no Brasil por conta de doenças contagiosas. De acordo com Liane Maria Bertucci-Martins (BERTUCCI-MARTINS, 2003), houve também o fechamento das escolas durante a crise da gripe espanhola em outubro de 1918 em São Paulo. Na época, o então presidente do Brasil, Delfim Moreira, aprovou todos os alunos. Segundo o relatório *Education at a Glance 2021* (OECD, 2021) da [Organização para a Cooperação e Desenvolvimento \(OCDE\)](#), o Brasil foi o país que mais tempo ficou com as escolas de nível infantil e fundamental fechadas, em 2020, entre 38 países. Esses países incluem os 37 membros da [OCDE](#), além do parceiro da [OCDE](#), a Federação Russa.

A [MDE](#) é uma área que se concentra na análise de dados educacionais, utilizando técnicas e ferramentas de mineração de dados, aprendizado de máquina e análise estatística para descobrir dados e informações. O principal objetivo da [MDE](#) é entender a educação e, a partir daí, desenvolver métodos que facilitem as trajetórias acadêmicas (ANDRADE; RIGO; BARBOSA, 2021). A [MDE](#) está em constante evolução com o intuito de desenvolver

formas de explorar os tipos de dados exclusivos em ambientes educacionais e entender os alunos e seus ambientes de aprendizado de maneira mais eficaz (KABAKCHIEVA, 2012). Existem várias tarefas na área de mineração de dados, como: regressão, classificação, agrupamento, regras de associações, correlações, dentre outras (BAKER, 2010).

No Brasil, o principal exame do ensino básico é o Exame Nacional do Ensino Médio (ENEM), o qual é feito com o intuito de avaliar os alunos que estão no último ano do ensino médio. Vale ressaltar que o exame é responsável por avaliar a educação básica, e utilizado para ingresso nas universidades públicas federais e estaduais. Sua nota pode ser também utilizada em instituições particulares de nível superior como forma de ingresso e até bolsas de estudo.

Entendendo a importância do exame do ENEM para a educação e por ser uma avaliação de desempenho de nível federal e possuir resultados com maior confiabilidade, para essa dissertação de mestrado, foram escolhidos os microdados do ENEM para entender o impacto da pandemia na educação básica e no principal exame da educação brasileira. Neste trabalho de mestrado busca-se minerar dados desse exame antes e durante a pandemia para entender o impacto da pandemia na educação básica do Brasil.

## 1.2 Caracterização do Problema

Com a chegada da pandemia, os órgãos responsáveis pela educação buscaram respostas sobre como agir e quais decisões tomar. No Brasil, o Ministério da Educação (MEC), órgão responsável pela gestão e qualidade da educação no País, buscava elaborar portarias para guiar os estados e municípios. O Brasil já vem sofrendo com problemas na educação mesmo antes da pandemia (PIETRA *et al.*, 2018). O Programa Internacional de Avaliação de Alunos (Programa Internacional de Avaliação de Alunos (PISA)) é um programa criado pela OCDE que visa avaliar a educação de 80 países e entender pontos que podem ser melhorados. Na edição de 2018 do PISA, aproximadamente 50% dos alunos brasileiros não atingiram a proficiência mínima na área de leitura. O país ficou em 57º lugar em leitura, 66º em ciências e em 70º em matemática conforme o relatório técnico (INEP, 2019).

Como visto, o Brasil apresenta baixos índices educacionais, aliados a baixos desempenhos, comparados a outros países. No período pandêmico, o Brasil ficou cerca de 178 dias com escolas fechadas, em 2020. Esses números foram divulgados no relatório *Education at a Glance 2021* (OECD, 2021). Além disso, o relatório informou que dentre os países analisados, o Brasil foi um dos países que não aumentou o orçamento para a Educação em 2020.

Mesmo durante o período de pandemia da COVID, as edições do ENEM continuaram acontecendo anualmente. Apesar do fechamento das escolas, orçamentos não

adequados no período, o exame de 2020 e 2021 continuou a ser realizado. Desta forma, o problema desta dissertação de mestrado é analisar o impacto da pandemia da COVID-19 no desempenho dos alunos no exame, buscando identificar possíveis mudanças ocorridas neste contexto. Nesse sentido, com a utilização de técnicas da área de Ciência da Computação, usando MDE, pretende-se investigar os impactos causados durante o período pandêmico, a fim de compreender melhor os desafios enfrentados pela educação básica durante a pandemia.

### 1.3 Hipótese de Pesquisa

A pandemia da COVID-19 pode ter impactado significativamente o desempenho dos participantes no ENEM, devido às mudanças na rotina escolar, na modalidade de ensino e no acesso à preparação para o exame. Neste sentido, a hipótese deste trabalho de mestrado é que, ao analisar diversas variáveis relacionadas aos participantes e utilizando de diferentes técnicas de MDE com as bases de dados do Instituto Nacional de Estudos Pedagógicos Anísio Teixeira (INEP), especificamente com os microdados do ENEM, pode-se indicar diferenças significativas no desempenho dos participantes e apontar possíveis consequências da pandemia para a educação á nível médio no Brasil.

### 1.4 Relevância do Trabalho

A educação é um dos pilares fundamentais de uma sociedade. Por meio dela, as pessoas adquirem conhecimentos e desenvolvem habilidades que permitem uma participação ativa na sociedade e uma melhor compreensão do mundo que as cerca. A educação é um dos componentes mais importantes no crescimento de qualquer país. Países com altas taxas de desenvolvimento humano, também, possuem altos índices educacionais e investimentos na qualidade da educação.

Quando se trata de estrutura educacional, a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) criou quatro pilares que podem servir de base, são eles: aprender a conhecer, aprender a fazer, aprender a conviver e aprender a ser. Além disso, outros fatores impactam diretamente na educação e na vida dos alunos. Fatores econômicos e sociais, por exemplo, impactam diretamente na qualidade da educação. Diversos estudos como o de Vasconcelos, et al. (VASCONCELOS *et al.*, 2021) e o de Pires (PIRES, 2015) já mostraram que a escolaridade dos pais, renda familiar, investimentos em infraestrutura das instituições de ensino são fatores, que estão correlacionados com o desempenho dos estudantes.

O Brasil é um país que possui altas taxas de analfabetismo e desempenho baixo em provas internacionais (MARINHO; BOCHEMBUZIO, 2022). Diferentemente da maioria

dos países da [OCDE](#), o Brasil não aumentou o investimento em educação durante o período da pandemia da COVID-19. Isto gera um alerta acerca da possível diminuição na qualidade da educação do Brasil, visto que o aumento em investimento na educação foi benéfico para outros países. Devido a essas adversidades, como falta de investimento, mudanças nas rotinas escolares e dificuldades causadas pela pandemia da COVID-19, é necessário entender os possíveis impactos causados na educação básica e como isso afeta o principal exame do ensino médio brasileiro.

## 1.5 Objetivos

Esse trabalho de mestrado tem o objetivo geral de minerar dados educacionais dos anos de 2017, 2018, 2019, 2020 e 2021 do [ENEM](#) para entender os impactos da pandemia da COVID-19 no exame e no ensino médio brasileiro nesse período.

Para tanto, consideram-se os seguintes objetivos específicos:

- Realizar análise exploratória dos dados do [ENEM](#) da educação básica brasileira dos anos de 2017, 2018, 2019, 2020 e 2021;
- Aplicar técnicas de mineração de dados para entender a situação educacional dos participantes do [ENEM](#) com a chegada da pandemia;
- Comparar as características como escolaridade dos pais, renda, tipo de escola, dentre outras e entender as relações nos dados educacionais;
- Identificar o impacto da pandemia da COVID-19 no desempenho dos participantes e nas características do [ENEM](#).

## 1.6 Organização do Trabalho

O restante desta dissertação de mestrado está organizada da seguinte forma:

- O **capítulo 2** apresenta a fundamentação teórica, apresentando os conceitos teóricos sobre a COVID-19, [ENEM](#) e [MDE](#), e que serviram de base para a elaboração desse estudo.
- O **capítulo 3** apresenta os trabalhos relacionados na área de estudo em questão, com o intuito de apresentar as contribuições e situar este trabalho em relação aos demais trabalhos já publicados.
- O **capítulo 4** apresenta a metodologia utilizada para a coleta, análise de dados, ferramentas utilizadas, descrevendo as etapas da pesquisa e os métodos adotados.

- No **capítulo 5** são descritos os resultados do trabalho, apresentando de forma individual os resultados de cada uma das questões de pesquisa.
- O **capítulo 6** contém as discussões sobre as questões de pesquisa, assim como os principais achados, limitações e verificação dos critérios de sucesso.
- O **capítulo 7** são descritas as considerações finais, assim como as contribuições deste trabalho para a ciência e as publicações geradas a partir dele.

## 2 Fundamentação Teórica

Neste capítulo, são apresentados os principais conceitos que servem de base para o restante do trabalho, visando fornecer uma base teórica para a análise e interpretação dos resultados da pesquisa. Este capítulo apresenta conceitos acerca da COVID-19, educação básica brasileira, como funciona o [ENEM](#) e todo seu processo até se tornar o principal exame de desempenho da educação, e o conceito de [MDE](#).

### 2.1 COVID-19 e a Educação Básica Brasileira

A COVID-19 é um vírus que eclodiu em Wuhan, província de Hubei, China. Apesar da infecção respiratória ficar conhecida recentemente, o vírus COVID surgiu de 1960 ([KHAN \*et al.\*, 2021](#)). A análise da sequência de várias espécies de coronavírus mostra que a doença é um vírus recombinante entre o coronavírus de morcego e o coronavírus de origem desconhecida, que se suspeita ser o pangolim ([CHAKRABORTY; MAITY, 2020](#)).

A COVID-19 pertence à família *Coronaviridae* que, por sua vez, é parte da ordem *Nidovirales*. As principais vias de transmissão da COVID-19 são de pessoas para pessoas por meio de gotículas respiratórias e contato pele a pele. Medidas clássicas de saúde pública, incluindo isolamento, quarentena, distanciamento social e contenção comunitária, foram usadas para conter a pandemia dessa doença respiratória ([ZHAI \*et al.\*, 2020](#)).

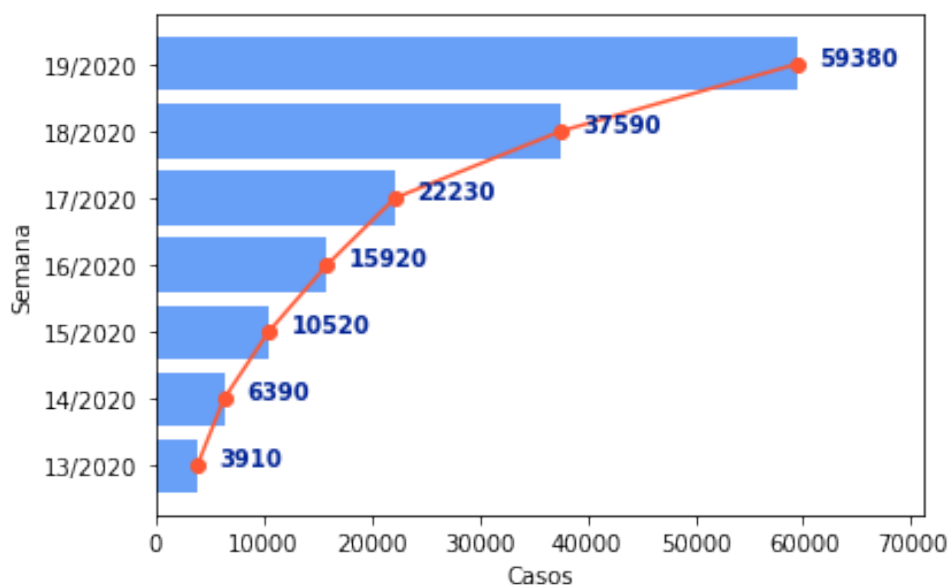
Até 31 de janeiro de 2023, foram identificados cerca de 670 milhões de casos da COVID-19 e 6.834.246 milhões de mortos no mundo. Em março de 2020, foi decretado pela [OMS](#) que a COVID-19 deixava de ser uma epidemia, ou seja, o aumento no número de casos da doença em diversas regiões, estados ou cidades, passou a atingir um nível mundial, se tornando assim uma pandemia. Depois do Ebola, H1N1, Zika Vírus, dentre outros, esta foi a sexta vez que a OMS declarava uma Emergência de Saúde Pública de Importância Internacional (ESPII).

As medidas de distanciamento social recomendadas pela OMS, como isolamento domiciliar de pessoas com sintomas da COVID-19, quarentenas, e uso de máscaras na população em geral, e implementadas na maioria dos países, resultaram também no fechamento de escolas públicas e particulares e na aulas em todos os níveis educacionais ([VIEIRA; SILVA, 2020](#)). O estudo ([ROBBINS, 2020](#)) mostrou que em 2020 cerca de 1,5 bilhão de estudantes em todo o mundo foram afetados pelo fechamento de escolas devido à COVID-19. Muitos países enfrentaram desafios em garantir que os estudantes tivessem acesso a recursos de qualidade para educação à distância. Também houve preocupações sobre a perda de aprendizado para os estudantes durante este período.

A COVID-19 também gerou um impacto econômico significativo em muitos países, comércios, bares, restaurantes ou qualquer negócio que não fosse considerado essencial teve que fechar suas portas por um determinado período, o que afetou a educação diretamente de várias maneiras. Por exemplo, alguns países tiveram que diminuir os orçamentos da educação para atender às necessidades econômicas mais prementes, como as crises de saúde, a contratação em massa de profissionais essenciais, a compra de insumos médicos, o que pode ter provocado um impacto negativo no acesso e na qualidade da educação. Além disso, muitos professores e coordenadores enfrentaram demissões ou cortes salariais devido à crise econômica, o que também pode ter impactado negativamente na educação. Alguns países também tiveram que tentar garantir que os alunos tivessem acesso a recursos de apoio, como refeições e equipamentos tecnológicos, internet, dentre outros, recursos que garantissem o aprendizado durante a pandemia (ELFIRDOUSSI *et al.*, 2020).

No Brasil, a doença se espalhou rapidamente, conforme apresentado na Figura 1. Esse avanço rápido no contágio levou a uma série de medidas de isolamento social pelo governo brasileiro. Escolas, restaurantes, estabelecimentos tiveram que se adequar às medidas, causando um forte impacto no setor econômico do país.

Figura 1 – Quantidade de casos de COVID-19 por região durante a décima terceira e a décima nona semana de 2020.



Fonte: Elaborado pelo autor (2020)

A pandemia de COVID-19 trouxe desafios e mudanças significativas para a educação no Brasil. Com o fechamento das escolas e universidades para evitar a disseminação do vírus, os estudantes se viram em situação de aprendizado à distância, o que pode ter sido um desafio para aqueles que não tinham acesso a equipamentos computacionais ou à internet com boas condições para uso. Além disso, professores tiveram que se adaptar rapidamente ao ensino online e encontrar novas formas de se comunicar e interagir com os



alunos. A falta de um planejamento adequado para adotar metodologias de ensino remoto ou à distância e a diminuição nos investimentos em infraestrutura educacional também foram um dos problemas, deixando estudantes e professores sem o planejamento correto para exercer as atividades educacionais (BARRETO; ROCHA, 2020).

Conforme a Portaria n.º 343 de 17 de março de 2020, o MEC determinou a substituição das aulas presenciais por aulas em meios digitais durante a pandemia. Em 28 de abril de 2020, o Conselho Nacional de Educação (CNE) emitiu parecer favorável à possibilidade de realização de atividades letivas não presenciais para cumprimento do ano letivo, o qual foi aprovado pelo MEC, que expediu portaria em 29 de maio de 2020 (MARTINS; ALMEIDA, 2020). Sistemas educacionais, escolas, professores, famílias e alunos tiveram que se adaptar rapidamente às aulas remotas. Sendo assim, os métodos de ensino à distância, começaram a ganhar mais força no contexto escolar.

No Brasil, a educação online é um conceito amplo e multifacetado (SANTANA; SALES, 2020). Uma educação de qualidade que quebre paradigmas e atenda à demanda por conteúdos inovadores para uma geração de alunos cada vez mais integrada à sociedade do conhecimento e da informação precisa ser cada vez mais estudada pelos pesquisadores e apoiada pelos governos (CASTRO; AGUILAR, 2014). Entretanto, vale ressaltar que uma experiência de aprendizado online bem projetada é muito diferente do ensino à distância que as instituições de ensino adotaram em resposta à crise da COVID-19. O modo repentino como a pandemia se instaurou fez com que os órgãos responsáveis pela educação não se planejassem de forma adequada e sem o investimento necessário para qual a demanda vinha. Um modelo de aprendizado online deve ser planejado e para atender todos os públicos, investindo em recursos computacionais, atendimentos personalizados e processos baseados em oferecer uma educação de qualidade para os alunos.

A segunda edição da pesquisa de resposta educacional à COVID-19 (INEP, 2021b) mostra que cerca de 99,7% das escolas brasileiras adotaram medidas preventivas e de controle da COVID-19 para o retorno das atividades presenciais em 2021. Em 2020, quase todas as escolas suspenderam as atividades presenciais no início da pandemia (99,3%). Cerca de 82,6% das escolas utilizaram o ensino presencial e/ou híbrido no ano de 2021.

O artigo de Vieira e Silva (SILVA, 2020) produziram uma *Revisão Sistemática de Literatura (RSL)* sobre as primeiras contribuições científicas brasileiras e portuguesas sobre os efeitos e desafios da pandemia do COVID-19 na educação durante o período pandêmico. Os autores mostram que o maior desafio para o Brasil durante a pandemia foi a adequação ao cenário pandêmico e às medidas impostas pelas autoridades na adoção de metodologias remotas, uso de plataformas online e sistemas de gestão acadêmica.

O censo escolar de 2020 realizado pelo INEP mostrou diversas fragilidades no âmbito da infraestrutura das escolas. No âmbito das escolas com dependência administrativa federal, cerca de 85% possuíam internet para o ensino e aprendizagem no ensino fundamental. Já

as escolas municipais, apenas 33,7% possuíam internet para o ensino e aprendizagem e entre as escolas estaduais do ensino fundamental, 71,5% possuíam internet para o ensino e aprendizagem. Os recursos tecnológicos ainda eram maiores para escolas de nível médio: 98% das escolas federais de ensino médio possuem internet para ensino e aprendizagem; as escolas municipais e estaduais de ensino médio possuem 46,4% e 64,6% respectivamente. Portanto, as escolas ainda não estavam preparadas para receber a pandemia (INEP, 2021a).

## 2.2 ENEM

Desde o surgimento, em 1998, o ENEM tem crescido cada vez mais, atestando sua importância no cenário nacional. O ENEM foi criado pelo Governo Federal para ser um instrumento auxiliar na avaliação do desempenho dos alunos ao final do ensino médio. Em 20 de agosto de 1998, o número de inscritos no ENEM era de 157.221 e o número de presentes naquela edição chegou a 115.575. Por mais de uma década, o exame foi usado exclusivamente para avaliar as competências e habilidades dos graduados do ensino médio, mas não para selecioná-los para o ensino superior (SILVEIRA; BARBOSA; SILVA, 2015).

O ENEM foi criado pelo governo federal, com o intuito de modernizar e reformar o sistema educacional brasileiro (MELO, 2012). Antes da criação do ENEM, o ingresso nas universidades públicas era feito por meio de vestibular aplicado pelas próprias instituições de ensino superior. O ENEM visa proporcionar escolhas mais justas e democráticas para o acesso ao ensino superior, visto que a prova é aplicada em contexto nacional e assim também poderia diminuir a quantidade de fraudes.

Além de ser um mecanismo de seleção para ingresso em universidades públicas, o ENEM também é utilizado como forma de avaliação do ensino médio no país. As notas obtidas pelos alunos são usadas para medir o desempenho dos alunos e da escola, e os resultados dos testes são divulgados para toda a sociedade através do Plano de Dados Abertos do INEP. Até o ano de 2019, o INEP disponibilizava os dados identificadores das escolas e dos alunos, sendo possível gerar um ranking das escolas. Devido à chegada da Lei Geral de Proteção de Dados (LGPD), os dados passaram a ser ocultados e não é mais possível gerar esse ranking.

A Tabela 1 mostra um breve histórico do ENEM e as mudanças que ocorreram ao longo dos 25 anos de realização da prova.

Ao longo dos anos, o ENEM passou a se aliar com outros programas e políticas de apoio e incentivo a educação. Programas como PROUNI, Fundo de Financiamento ao Estudante do Ensino Superior (FIES), Reestruturação e Expansão das Universidades Federais (REUNI) e SISU foram os principais aliados do exame, tornando-a referência nacional (TRAVITZKI, 2013). Segundo Viggiano e Mattos (VIGGIANO; MATTOS, 2019), com a mudança ocasionada em 2009, o exame, que antes servia apenas como modelo de

Tabela 1 – Histórico do ENEM.

Ano	Alterações e Eventos Importantes
1998	Criação do exame
2001	O exame passou a receber inscrições pela internet e o número de municípios subiu de 184 para 277.
2004	Utilização da nota do ENEM no Programa Universidade Para Todos ( <a href="#">Programa Universidade Para Todos (PROUNI)</a> ) para concessão de bolsas para os participantes.
2009	A criação do Sistema de Seleção Unificada ( <a href="#">Sistema de Seleção Unificada (SISU)</a> ) faz com que o ENEM seja o principal meio de egresso nas universidades.
2013	Todas as universidades federais do País usam a nota do ENEM como critério de seleção.
2020	Aplicação do exame de forma digital pela primeira vez.

avaliação, mudou para o fenômeno de “vestibularização” do [ENEM](#). Desta forma, o exame passou a ter 180 questões objetivas: 45 questões da área de Ciências Humanas e suas Tecnologias, que conta com as disciplinas de História, Geografia, Filosofia e Sociologia; 45 questões da área Ciências Naturais e suas Tecnologias, composta pelas disciplinas de Química, Física e Biologia; 45 questões da área de Linguagens, Códigos e suas Tecnologias, que constam com as disciplinas de Língua Portuguesa, Literatura, Língua Estrangeira (Inglês ou Espanhol), Artes, Educação Física e Tecnologias de Informação e Comunicação; 45 questões da área de Matemática e suas Tecnologias; e a Redação.

De acordo com ([TRAVITZKI, 2013](#)), não somente as características da prova tiveram alterações, os objetivos e competências avaliados também mudaram conforme a Tabela 2.

1998 - 2008	2009 - 2022
Avaliar a preparação para a cidadania	Avaliar a preparação para o Ensino Superior
5 competências e 21 habilidades	4 áreas de conhecimento e 30 competências
1 prova de 63 itens e 1 redação	4 provas de 45 itens e 1 redação
Teoria Clássica dos Testes	Teoria da Resposta ao Item

Tabela 2 – Principais mudanças realizadas na prova em 2009 ([TRAVITZKI, 2013](#)).

Ainda de acordo com ([TRAVITZKI, 2013](#)), a mudança da forma de análise dos resultados é uma das mudanças principais do novo [ENEM](#). Antes a prova não possuía áreas de conhecimento, ela era composta por 3 itens: fácil, médio e difícil. Cada item contemplava 21 habilidades, somando 63 questões. A forma de análise era a teoria clássica dos testes, ou seja, a nota era basicamente a soma de quantas questões foram marcadas de forma correta. Depois disso, a nota era somada com a nota da redação.

As evoluções educacionais e a necessidade de uma prova que englobasse mais habilidades e competências, fez com que o exame passasse por mudanças. O ENEM passou a ter 120 habilidades, sendo 30 para cada uma das quatro grandes áreas que compõem o exame. A divisão da prova passou por 4 áreas de conhecimentos que possuíam eixos cognitivos baseados no modelo anterior do exame, apresentados na Tabela 3. No atual modelo, a nota é calculada pela soma da pontuação obtida nas provas das 4 áreas de conhecimento e a da redação, dividido por cinco.

O método de análise dos resultados mudou e passou a utilizar a Teoria de resposta ao item (TRI). O ENEM utiliza um modelo de TRI específico, conhecido como Teoria de Resposta ao Item de Três Parâmetro. Esse modelo de avaliação é baseado em uma escala de proficiência, que mede o grau de habilidade dos estudantes em cada uma das competências avaliadas. Essa escala permite comparar o desempenho dos estudantes em diferentes edições da prova, bem como avaliar o impacto de políticas públicas voltadas para a melhoria da educação. O nível das questões é definido depois que a prova já foi feita, e a pontuação é aplicada conforme a taxa de acertos e erros que cada questão teve (ANDRADE; TAVARES; VALLE, 2000), ou seja, um participante que erra as questões mais fáceis e acerta algumas difíceis de forma aleatória, pela falta de coerência, o resultado dele acaba sendo menor do que um participante que acerta as fáceis e medianas e erra as difíceis, pois, segundo a TRI, o participante teve mais coerência, e então aquelas respostas não são consideradas “chutes”.

Tabela 3 – Eixos Cognitivos do ENEM (CAMPOS, 2015).

Dominar linguagens	Compreensão das normas cultas do português, usar a linguagem da matemática, artes e ciências, bem como espanhol e inglês.
Compreender fenômenos	Construir e aplicar conceitos de diversas áreas do conhecimento para compreender fenômenos naturais, geográficos, históricos, produção tecnológica e expressão artística
Enfrentar situações-problema	Selecionar, ordenar, comparar e interpretar dados e informações apresentados de diferentes formas de modo a enfrentar problemas e tomar decisões.
Construir argumentação	Vincular as informações apresentadas de diferentes maneiras com o conhecimento disponível em uma situação específica para construir argumentos coerentes.
Elaborar propostas	Utilizar os conhecimentos aprendidos na escola para elaborar propostas de intervenções solidárias na realidade, respeitar os valores humanos e consideração da diversidade sociocultural.

Outro grande evento na história do exame foi a edição digital da prova. Até 2019, o exame era aplicado exclusivamente em formato impressão. Antes mesmo da pandemia,

em 2015, o governo já estudava a aplicação do exame de forma digital. A mudança para o formato digital foi anunciada pelo MEC em 2019 e segundo o próprio ministério, o exame passaria a ser completamente digital no ano de 2026. O ENEM em formato digital é aplicado como uma prova online, em que o aluno pode acessar as questões e preencher as respostas diretamente no computador. De acordo com (JUNIOR, 2021), a realização do ENEM de forma digital pode resultar em vários benefícios: celeridade na correção das questões e da redação, diminuição de gastos com impressões, aumento na segurança contra vazamentos, tornando-a menos vulnerável a fraudes.

## 2.3 Mineração de Dados Educacionais

Mineração de dados é o processo de descoberta de padrões, relacionamentos e *insights* em grandes conjuntos de dados usando estatística, inteligência artificial e aprendizado de máquina (GHAVAMI, 2019). Essas informações podem ser utilizadas para diversos fins, como a tomada de decisões empresariais, a identificação de padrões em grandes conjuntos de dados, a previsão de tendências e a detecção de problemas ou oportunidades (AMARAL, 2016).

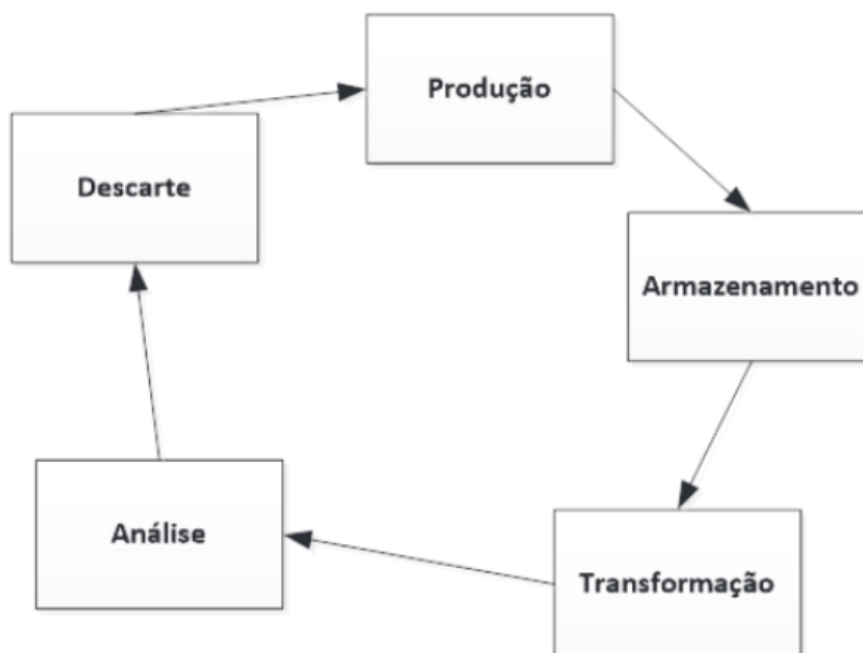
De acordo com Aher e Lobo (AHER; L.M.R.J., 2011), a mineração de dados pode ser usada em ambientes educacionais com o intuito de prever as taxas de evasão, a relação entre as notas dos testes de admissão e o sucesso dos alunos, prever o desempenho acadêmico dos alunos, encontrar evidências de fatores que afetam o desempenho e melhorar a gestão destinada à política educacional.

A Figura 2 mostra o ciclo de vida dos dados, exibindo as etapas que um dado pode passar de acordo com sua finalidade. É preciso analisar o que fazer com os dados antes de estudá-los, portanto, entender o ciclo de vida dos dados é um processo essencial para definir os objetivos de um projeto de mineração de dados.

A mineração de dados é uma área que está em constante evolução, e pode ser aplicada a uma variedade de domínios, incluindo finanças, saúde, comércio, transporte e muitos outros (SUNHARE; CHOWDHARY; CHATTOPADHYAY, 2022). Essa evolução é possível graças ao crescente volume de dados gerados pelas atividades humanas, especialmente na era da tecnologia da informação. Esses conjuntos de dados são geralmente gerados por fontes diversas, como sensores, dispositivos móveis, redes sociais e transações comerciais, entre outras. Esse grande volume de dados é conhecido com *Big Data* ( TSAI *et al.*, 2015).

Existem várias técnicas utilizadas na mineração de dados, incluindo aquelas para a realização de tarefas de regressão, classificação, agrupamento, dentre outras. Cada técnica é adequada para um tipo específico de problema ou objetivo, e é importante escolher a abordagem certa para o trabalho em questão. Por exemplo, quando se busca prever um

Figura 2 – Etapas do ciclo de vida dos dados



Fonte: [Amaral \(2016\)](#).

valor numérico, se utiliza algoritmos de regressão, mas quando se está querendo prever uma variável categórica, se usa técnicas de classificação. A mineração de dados pode requerer um poder de processamento maior, pois, como dito anteriormente, os conjuntos de dados podem ser enormes e exigir muitos cálculos para serem analisados. Além disso, é importante ter um bom entendimento dos dados e do problema que se quer resolver, pois isso pode afetar a escolha das técnicas e a interpretação dos resultados ([GOLDSCHMIDT; PASSOS, 2005](#)).

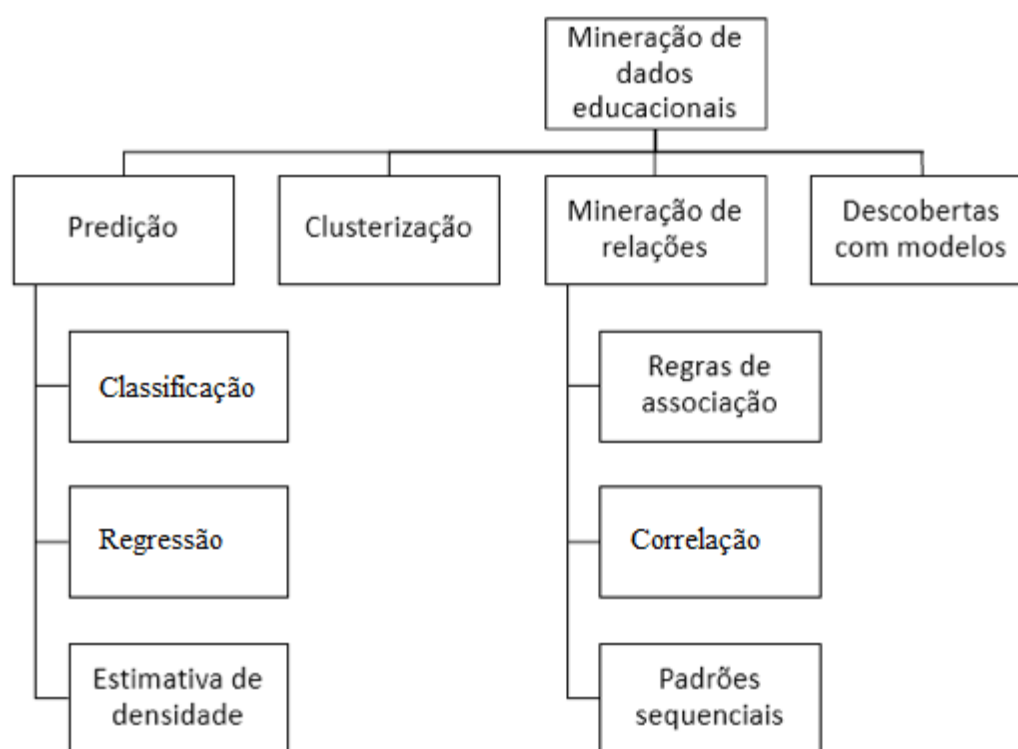
A [MDE](#) transforma dados brutos dos sistemas educacionais em informações pertinentes, com o potencial de ter um maior impacto na pesquisa e na prática educacional ([DUTT et al., 2015](#)). A [MDE](#) pode ser usada, por exemplo, para o melhor entendimento de como os alunos estão lidando com um novo sistema acadêmico, assim como, para prever se um aluno está mais propenso a evadir a partir determinadas características.

A área de [MDE](#) é um campo de estudo que usa técnicas de mineração de dados para entender os dados provindos de sistemas educacionais. Métodos e ferramentas de *e-learning* interativos oferecem oportunidades para coletar e examinar os dados dos alunos, professores e todos os atores envolvidos em um ambiente educacional, identificar padrões e tendências nos dados, e realizar descobertas e testar hipóteses sobre como os alunos aprendem, por exemplo ([PRABHA; SHANAVAS, 2014](#)).

A Figura 3 mostra como as tarefas de mineração de dados são divididas por Santos et al. ([SANTOS; SILVEIRA; LECHUGO, 2017](#)). As tarefas de predição são utilizadas

quando se deseja prever um resultado ou evento futuro com base em dados anteriores. As tarefas de predição são divididas em classificação, regressão e estimativa de densidade. A classificação visa identificar a qual categoria pertence uma determinada nova amostra, para a qual o valor da variável de destino é do tipo categórico. As tarefas de regressão são semelhantes às tarefas de classificação, entretanto a diferença é que o atributo de destino assume um valor numérico. No caso da estimativa de densidade, a variável de destino é uma função de densidade de probabilidade.

Figura 3 – Tarefas de MDE



Fonte: SANTOS, Silveira e LECHUGO (2017)

A clusterização é um processo de agrupamento de dados em conjuntos (ou clusters) de modo que os elementos de um mesmo cluster sejam mais similares entre si do que aos elementos de outros clusters (BERKHIN, 2006). A aplicação de técnicas de agrupamento são geralmente usadas quando se deseja agrupar objetos semelhantes em conjuntos, com base em suas características ou atributos. Alguns tipos de algoritmos de agrupamento são: agrupamento hierárquico, em que se agrupa dados em um número de grupos hierárquicos com base na distância entre eles; DBSCAN, em que o agrupamento é baseado na densidade alta e baixa entre os dados; e K-Means, em que o agrupamento é baseado na distância euclidiana entre os pontos de dados.

A regra de associação é uma técnica de aprendizado de máquina que busca identificar padrões e relações entre diferentes variáveis em um conjunto de dados (BAKER, 2010).

As regras de associação são baseadas no conceito de suporte e confiança. O suporte de uma regra é a frequência com que os itens da regra aparecem juntos no conjunto de dados (RIBEIRO; VIEIRA; TRAINA, 2005). Por exemplo, uma regra de associação pode ser encontrada na forma de: variável 1 - o aluno está frustrado, pois a meta dele é maior que a capacidade de ganhar conhecimento, logo, a variável 2 vai ser - o aluno frequentemente pede ajuda. Ou seja, se algum conjunto de valores de variáveis for encontrado, outra variável geralmente terá um valor específico, neste caso a variável com o valor específico é a variável 2.

A correlação é um conceito importante em mineração de dados, que visa encontrar correlações lineares ou não, positivas ou negativas (BAKER, 2010). As medidas de correlação são usadas para quantificar o grau de associação ou relação entre variáveis em um conjunto de dados. Ela é medida por um valor chamado de coeficiente de correlação, que varia de -1 a 1. Quanto mais próximo de 1, maior a correlação positiva entre as variáveis. As três principais medidas de correlação são a correlação de Pearson, de Kendall e de Spearman. A correlação de Pearson é uma medida de correlação linear que mede a relação entre variáveis contínuas normalmente distribuídas. A correlação de Kendall e a de Spearman são duas medidas de correlação não paramétricas usadas quando as variáveis envolvidas são ordinais ou de classificação. A de Kendall mede a concordância ou discordância entre as variáveis e a de Spearman mede o grau de associação entre as variáveis.

A descoberta com modelos é tratada como um método que se soma as tarefas de predição, agrupamento e mineração de relações. Essas tarefas geram modelos de aprendizado de máquina que pode servir para a descoberta das relações entre os comportamentos dos estudantes e as variáveis que o cercam. A aplicação de técnicas de mineração de dados para extrair conhecimento de sistemas educacionais vem sendo tratada por (ARRUABARRENA *et al.*, 2002) como uma forma de avaliação formativa com propósito de melhorar continuamente o processo de evolução do sistema educacional. O conhecimento descoberto através dessas técnicas pode ser utilizado pelas pessoas envolvidas, perpassando desde os alunos, professores, gestão educacional, e até mesmo os responsáveis pelos estudantes.

## 2.4 Conclusão

Neste capítulo foi visto como a pandemia da COVID-19 se instaurou pelo mundo e como a doença gerou uma série de mudanças no cenário educacional brasileiro. Foi apresentado o ENEM, sua história e características relevantes para esse trabalho de mestrado. Também apresentaram-se os conceitos ligados a MDE. Desta forma, o capítulo apresenta todo o referencial teórico usado para o desenvolvimento do restante desta dissertação.



## 3 Trabalhos Relacionados

O objetivo deste capítulo é fornecer uma visão abrangente das pesquisas já realizadas sobre o tema desta pesquisa de mestrado. O capítulo está dividido em três seções. Na primeira seção, são apresentados os trabalhos que utilizaram os dados do [ENEM](#) para minerar dados. A segunda seção apresenta os trabalhos relacionados que buscaram através de [MDE](#) entender a educação na pandemia. Por fim, na terceira seção, é apresentada uma discussão sobre os trabalhos relacionados e uma análise comparativa entre os trabalhos e essa dissertação de mestrado, a fim de evidenciar as contribuições desse estudo.

### 3.1 Mineração de Dados do ENEM

A revisão sistemática de ([SOARES et al., 2021](#)) foca nos trabalhos que utilizam técnicas de [MDE](#) 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](#). A [RSL](#) feita por Lima et al. ([LIMA et al., 2019](#)) visa identificar os tipos de análises que têm sido usadas em trabalhos que utilizam os dados do [Exame Nacional de Desempenho de Estudantes da Educação Superior \(ENADE\)](#) e do [ENEM](#). Foram selecionados 17 trabalhos que utilizavam dados do [ENEM](#) e apenas um deles utilizava técnicas de mineração de dados. A [RSL](#) mostrou que o número de trabalhos que usam o [ENADE](#) é maior. Para essa revisão foram selecionados 40 trabalhos com dados do [ENADE](#).

O trabalho de ([SILVA; MORINO; SATO, 2014](#)) utilizou os dados do [ENEM](#) do ano de 2010, usando uma tarefa de 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 estudo mostrou que a classe social, a escolaridade dos pais e a quantidade alta de pessoas que moram com o estudante são as características que mais causam diminuição no desempenho.

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. O estudo utilizou uma amostragem de 950 mil participantes e descobriu que aspectos como escolaridade do pai, internet em casa e classe social, tem indícios que afetam a média do estudante.

Já o artigo de ([NAKAZONE; 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. O estudo mostrou que a mudança da prova para acontecer em dois domingos

aumentou o número de ausentes. O estudo mostrou que as regiões sul e sudeste tem notas maiores quando comparado com outras regiões do país.

O artigo de (FRANCO *et al.*, 2020) faz uma análise através de mineração de dados para identificar os fatores que mais influenciam o ENEM. Para isso, eles analisaram 22 anos (1998-2019) do ENEM. Através da aplicação de algoritmos de seleção e classificação de fatores, os autores identificaram os 20 principais fatores que podem contribuir no desempenho dos estudantes no exame, sendo alguns deles: tipo de escola que cursou o ensino médio; ter computador e quantos; está frequentando um curso profissionalizante; e escolaridade da mãe.

Ainda em busca de analisar o desempenho dos estudantes, Alves et al. (ALVES; CECHINEL; QUEIROGA, 2018) visaram criar um modelo preditivo do desempenho das notas da prova de Matemática e suas Tecnologias do ENEM. Através do WEKA, uma ferramenta de mineração de dados que utiliza a linguagem JAVA, os autores utilizam os algoritmos Naive Bayes e J48 para prever as notas dos participantes. As variáveis de dependência administrativa e indicador de nível socioeconômico foram as que mais impactaram no resultado dos experimentos.

Outro artigo que busca prever o desempenho dos participantes no ENEM é o estudo de Alves et al. (ALVES *et al.*, 2018). Usando técnicas de mineração de dados e os algoritmos Naive Bayes e J48, os autores desenvolveram um modelo capaz de prever o desempenho dos estudantes com acurácia de 61.74% das amostras presentes na base de dados do ENEM de 2016.

Seguindo o objetivo de analisar outros fatores relevantes da prova, Silva et al. (SILVA *et al.*, 2020) buscaram, através de tarefas de mineração de dados, identificar as desigualdades sociais dos participantes do ENEM. No estudo, os autores usaram algoritmos de clusterização e regras de associação para encontrar as variáveis determinantes no desempenho dos participantes do exame. Por fim, através das variáveis relacionadas ao tipo de escola, raça, renda familiar e escolaridade da mãe, os autores dividiram os participantes em dois grupos, os que alcançaram nota acima de 585,6 e outro com participantes com notas abaixo de 585,6.

A aplicação de mineração de dados como ferramenta de descoberta de fatores associados ao desempenho dos alunos do ensino médio é o que propõe o artigo Araújo et al. (ARAÚJO; SILVA, 2020). Utilizando regras de associação com o algoritmo FP-Growth, os autores utilizaram os microdados dos anos de 2016, 2017 e 2018 com os dados do estado de Goiás. Para isso, usaram uma métrica de confiança. A educação pública aparece nas notas mais baixas e os participantes com notas médias mais baixas tiveram declaração de renda abaixo de dois salários mínimos.

## 3.2 Mineração de Dados Educacionais e COVID-19

Diversos estudos ao redor do mundo têm focado nos impactos da pandemia da COVID-19 na educação. Para evidenciar a importância do tema, o editorial de (REUGE *et al.*, 2021) apresenta uma edição especial proposta pela [United Nations Children’s 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 *et al.* (JUNIOR; MATOS; BORGES, 2021) realiza uma análise pelos algoritmos Self-organizing maps (SOM) e K-means para o experimento com os estudantes durante a pandemia da COVID-19. Para cada algoritmo, foi criado 3 grupos diferentes. Os resultados mostraram que os alunos que possuíam perfil com melhores condições de acesso à internet e infraestrutura eram mais engajados nas atividades propostas com a metodologia de [Educação a Distância \(EAD\)](#).

O estudo de Moraes *et al.* (MORAES; PERES; PEDREIRA, 2021) entendeu as relações das variáveis apresentadas nos microdados do [ENEM](#) durante a pandemia, com o intuito de avaliar o desempenho de alunos na área de conhecimento de matemática e suas tecnologias. Os resultados desse estudo mostraram que alunos provenientes de famílias de renda superior alta possuem notas superiores ao de alunos de renda inferior. Por fim, o estudo fala que a diferença financeira existente entre os estudantes tende a aumentar durante a pandemia, e ressalta que os governantes e gestores não intervindo e criando políticas públicas adequadas, essa disparidade pode aumentar mais ainda.

Motivados pelos possíveis danos que a pandemia da COVID-19 poderia causar e utilizando uma amostragem de 49 alunos do curso de Ciências da Saúde durante um bloqueio das aulas de uma universidade da Espanha causado pelo COVID-19, Sáiz-Manzanares *et al.* (SAIZ-MANZANARES *et al.*, 2021) propuseram a ferramenta *UBUMonitor* para avaliar os danos da pandemia. Ela é uma aplicação que usa dados do *Moodle* para aplicar técnicas de [MDE](#) para monitorar e detectar alunos em risco de evasão. Realizando análises estatísticas com uso de visualização de dados com *boxplots* e *heatmaps*, e algoritmos de clusterização como K-means, fuzzy k-means e o DBSCAN, foi possível acompanhar os dados dos alunos que resultaram em uma taxa de evasão de 7% e uma elevada satisfação dos alunos com o processo de ensino-aprendizagem.

Santos *et al.* em (SANTOS *et al.*, 2021) usaram o Twitter através de técnicas de análise de discurso por [Processamento de Linguagem Natural \(PLN\)](#) e análise de sentimentos para entender os discursos publicados no Twitter e identificar se era um discurso positivo ou negativo em relação à educação, considerando o cenário pandêmico da

COVID-19. Foram identificadas palavras quando se fala de COVID e educação através do uso de uma nuvem de palavras. Através de um agrupamento de palavras no período de 13 a 18 de novembro de 2021, foram geradas as análises de sentimentos. Durante o período analisado, o estudo mostrou que o pessimismo foi o sentimento dominante ao buscar pelos termos educação e COVID-19.

### 3.3 Discussão

As áreas de pesquisa relacionadas à mineração de dados evoluíram bastante, o que possibilitou o crescente número de trabalhos usando mineração de dados para resolver problemas educacionais. Dentre os anos de 2006 a 2010, o número de trabalhos publicados sobre MDE era de apenas quinze no Brasil, segundo o estudo de Rodrigues et al. (GOMES, 2015). Já de 2011 até meados de 2014, o mesmo estudo mostra que foram publicados cerca de 50 artigos, evidenciando uma evolução significativa.

Conforme apresentado, já existem diversos trabalhos utilizando MDE 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 MDE de modo a comparar os anos de 2017, 2018, 2019, 2020 e 2021 do ENEM para compreender o impacto da pandemia da COVID-19. Portanto, a contribuição desta dissertação de mestrado é a realização de uma análise dos dados dos microdados do ENEM entre os anos anteriores à pandemia (2017, 2018 e 2019) e os anos durante a pandemia (2020 e 2021) para identificar se houve impacto na quantidade de inscritos, presentes e ausentes no exame, no desempenho dos estudantes em consideração a sua região, renda financeira, nos fatores relacionados a dados socioeconômicos, nas correlações destes fatores com a nota, nas áreas de conhecimento foram impactadas durante esse período e na características dos participantes que realizaram o exame.

## 4 Metodologia

Este capítulo apresenta, primeiramente, a metodologia [CRISP-DM](#), a qual foi utilizada nesta dissertação para analisar os dados. Em seguida, são apresentadas as questões de pesquisa, a motivação para cada uma delas e os critérios de sucesso. Será apresentado as etapas do [CRISP-DM](#) e como elas foram utilizadas nessa dissertação. O capítulo apresenta também as ferramentas utilizadas e como as etapas de preparação dos dados, visualização e modelagem dos dados foram feitas.

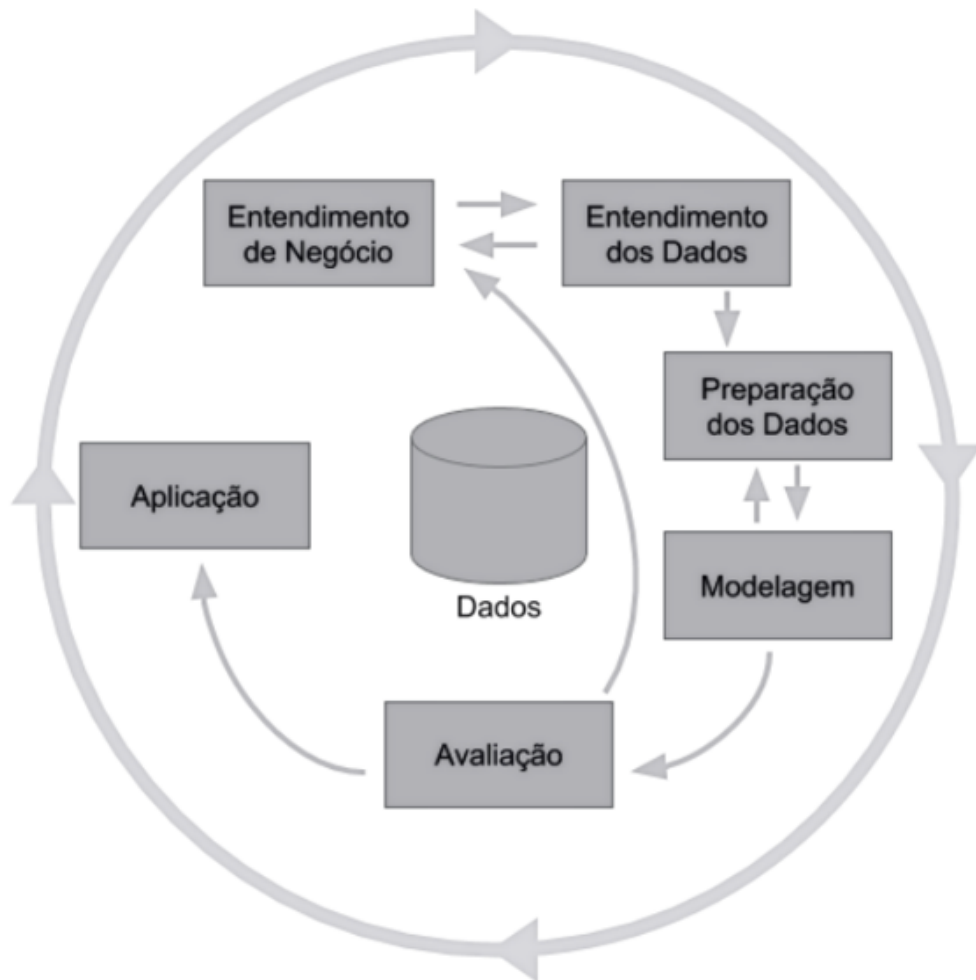
### 4.1 CRISP-DM

A metodologia [CRISP-DM](#) fornece um conjunto estruturado de etapas para a condução de um projeto de mineração de dados, sua abordagem iterativa, permite a revisão e aprimoramento do projeto em cada etapa, devido a essas características, o [CRISP-DM](#) foi escolhido como metodologia para essa dissertação. O [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; SANTOS, 2008](#); [SHEARER, 2000](#)).

A Figura 4 mostra o ciclo do [CRISP-DM](#). Na etapa de entendimento do negócio, o objetivo principal é entender o problema de negócios e as metas que devem ser alcançadas. A etapa de entendimento dos dados é quando a coleta dos dados são feitas e quando se encontra as primeiras informações sobre eles. Preparação de dados é a etapa responsável por fazer a limpeza, preparação e transformação dos dados que irão passar pelo processo de mineração. Na etapa de modelagem, são aplicadas técnicas de mineração de dados para construir modelos. Depois da construção dos modelos, é feita a etapa de avaliação. Por fim, a distribuição é a etapa em que os modelos são implantados no negócio. Vale ressaltar que todo o processo do [CRISP-DM](#) é iterativo e, a cada ciclo, é possível já extrair conhecimento dos dados.

O estudo de ([RAMOS \*et al.\*, 2020](#)) propõe a utilização de uma adaptação do metodologia [CRISP-DM](#) para uma abordagem educacional. Nesta adaptação, ao invés de usar uma metodologia voltada para a indústria, como diz o nome da metodologia, o autor sugere que a utilização dessa metodologia consegue auxiliar os pesquisadores da área da educação, indicando um caminho metodológico para ambientes educacionais. Na proposta feita pelo autor, as etapas permanecem as mesmas, entretanto, o ambiente onde se ocorre a distribuição do conhecimento é um ambiente educacional, sendo similar a este trabalho de mestrado.

Figura 4 – Ciclo da vida do CRISP-DM



Fonte: Adaptado de Shearer (2000)

## 4.2 Questões de Pesquisa e Critérios de Sucesso

Este estudo objetiva responder às seguintes *Questões de Pesquisa (QPs)*:

- (QP1) Qual o impacto da pandemia da COVID-19 na quantidade de inscritos presentes e ausentes no ENEM?
  - Motivação: Buscamos analisar se a pandemia fez com que os participantes deixassem de realizar o exame.
- (QP2) Qual o impacto da pandemia da COVID-19 no desempenho dos participantes, considerando o tipo de dependência administrativa da escola em que terminaram o ensino médio?
  - Motivação: O desempenho dos participantes provindos de escola pública é, historicamente, inferior aos participantes de escola particular. Então visamos

analisar as características dos alunos de escola pública em relação aos oriundos de escola particular.

- (QP3) Qual o impacto da pandemia da COVID-19 no desempenho dos participantes do ENEM, considerando a renda familiar?
  - Motivação: Os estudos prévios apresentados na Seção 3 mostram que os participantes com maior renda familiar possuíam maiores notas no ENEM. Levando isso em consideração, visamos analisar os impactos da pandemia para participantes considerando a renda familiar.
- (QP4) Quais são as áreas do conhecimento do ENEM mais impactadas durante o período da pandemia da COVID-19?
  - Motivação: Visamos identificar se e quais das 5 áreas do conhecimento do exame sofreram mais impacto durante a pandemia da COVID-19.
- (QP5) Quais as variáveis socioeconômicas que possuem maior influência com a nota média dos participantes antes e durante a pandemia da COVID-19?
  - Motivação: Buscamos descobrir quais das características socioeconômicas podem influenciar na nota dos participantes e analisar se essa correlação mudou durante o período da pandemia da COVID-19 no Brasil.
- (QP6) Qual o impacto da pandemia no desempenho dos participantes do ENEM para as diferentes regiões e os estados?
  - Motivação: Durante o período de pandemia, os estados tiveram que fazer economias e ajustar suas políticas educacionais. Nessa questão, objetivamos analisar se a pandemia afetou o desempenho dos participantes em relação aos estados e regiões.
- (QP7) Houve mudança nas características dos participantes do ENEM durante o período de pandemia de acordo com questões socioeconômicas?
  - Motivação: Nessa questão, visamos analisar as características de renda, acesso à tecnologia e escolaridade dos pais dos participantes do ENEM e se elas mudaram durante o período da pandemia da COVID-19.

Os *Critérios de Sucesso (CSs)* dessa pesquisa são listados abaixo:

- (CS1) Atingir os objetivos definidos para esse estudo e encontrar as respostas para as questões de pesquisa;

- (CS2) Identificar os impactos causados pela pandemia da COVID-19 para os estudantes da educação ao nível médio no [ENEM](#).
- (CS3) Publicar artigos científicos para realizar a fase de distribuição da metodologia [CRISP-DM](#), além da escrita desta dissertação, a qual relata todos os resultados em um único documento.

### 4.3 Entendimento do Negócio e dos Dados do ENEM

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. Foram utilizados cinco conjuntos de dados para esse trabalho: os microdados do [ENEM](#) dos anos de 2017, 2018, 2019, 2020 e 2021, ambos disponibilizados pelo [INEP](#) em seu portfólio de dados abertos ([INEP, 2023](#)).

O [ENEM](#) é uma avaliação criada e aplicada pelo Ministério da Educação para medir o desempenho de alunos que concluíram o ensino médio. O [ENEM](#) foi originalmente concebido como um exame final para avaliar os alunos ao final da educação básica. No entanto, nos últimos anos, ele tem sido visto como um teste de admissão para universidades públicas e privadas. Atualmente a nota do [ENEM](#) é composta através do domínio em diferentes áreas conforme detalhado no capítulo 2. Entretanto, não só os dados relacionados as provas podem ser interessantes. Durante o ato de inscrição, o estudante responde a um questionário com 25 questões contendo informações sobre seu nível socioeconômico, família, educação e trabalho, além dos dados dos participantes, contendo idade, sexo, ano de conclusão do ensino médio, tipo da escola (pública ou privada), dentre outras informações que podem gerar conhecimento.

Os microdados do [ENEM](#) fazem parte do conjunto de dados abertos do [INEP](#), estando disponíveis publicamente. Os microdados são compostos da maneira em que cada linha representa um participante inscrito no exame e as colunas são características de cada um deles.

A Tabela 4 mostra as variáveis utilizadas ao longo da dissertação e uma descrição resumida. As variáveis `NU_NOTA_CN`, `NU_NOTA_CH`, `NU_NOTA_LC`, `NU_NOTA_MT` e `NU_NOTA_REDACAO` são do tipo numérica, sendo utilizadas para avaliar desempenho. As variáveis `TP_DEPENDENCIA_ADM_ESC` e `SG_UF_PROVA` são do tipo categórica, sendo utilizadas para agrupar os participantes em relação ao seu desempenho. As variáveis `Q001` até `Q025` fazem parte do questionário socioeconômico. Das variáveis socioeconômicas, apenas a `Q005` que se trata da quantidade de pessoas morando na residência é do tipo numérica.



Tabela 4 – Descrição das variáveis utilizadas.

<i>Variável</i>	<i>Descrição</i>	<i>Tipo de variável</i>
NU_NOTA_CN	Nota da prova de Ciências da Natureza	Numérica
NU_NOTA_CH	Nota da prova de Ciências Humanas	Numérica
NU_NOTA_LC	Nota da prova de Linguagens e Códigos	Numérica
NU_NOTA_MT	Nota da prova de Matemática	Numérica
NU_NOTA_REDACAO	Nota da prova de redação	Numérica
TP_DEPENDENCIA_ADM_ESC	Tipo de dependência administrativa da escola de conclusão do ensino médio	Categórica
SG_UF_PROVA	Sigla da Unidade da Federação da aplicação da prova	Categórica
Q001	Qual a escolaridade do Pai?	Categórica
Q002	Qual a escolaridade da Mãe?	Categórica
Q003	Qual ocupação do seu Pai?	Categórica
Q004	Qual ocupação da sua Mãe?	Categórica
Q005	Qual a quantidade de moradores na sua residência?	Numérica
Q006	Qual a sua renda familiar?	Categórica
Q007	Em sua residência trabalha empregado(a) doméstico(a)?	Categórica
Q008	Na sua residência existe banheiro?	Categórica
Q009	Quantidade de quartos na sua residência?	Categórica
Q010	Na sua residência tem carro?	Categórica
Q011	Na sua residência tem motocicleta?	Categórica
Q012	Na sua residência tem geladeira?	Categórica
Q013	Na sua residência tem freezer?	Categórica
Q014	Na sua residência tem máquina de lavar roupa?	Categórica
Q015	Na sua residência tem máquina de secar roupa?	Categórica
Q016	Na sua residência tem forno micro-ondas?	Categórica
Q017	Na sua residência tem máquina de lavar louça?	Categórica
Q018	Na sua residência tem aspirador de pó?	Categórica
Q019	Na sua residência tem televisão em cores?	Categórica
Q020	Na sua residência tem aparelho de DVD?	Categórica
Q021	Na sua residência tem TV por assinatura?	Categórica
Q022	Na sua residência tem telefone celular?	Categórica
Q023	Na sua residência tem telefone fixo?	Categórica
Q024	Na sua residência tem computador?	Categórica
Q025	Na sua residência tem acesso à Internet?	Categórica

## 4.4 Ferramentas Utilizadas

Para o entendimento dos dados e para avançar até a etapa de preparação dos dados, é preciso fazer o planejamento de quais tecnologias serão usadas. Esta seção apresenta as

principais ferramentas utilizadas durante a construção dessa pesquisa de mestrado.

*Python* é uma linguagem de programação de alto nível, interpretada e de código aberto. Criada em 1989 por Guido van Rossum, essa linguagem tem sido amplamente utilizada em diversas áreas, como desenvolvimento web, análise de dados, inteligência artificial e mineração de dados (SOUSA *et al.*, 2020). O *Python* ganhou popularidade como uma linguagem amigável para iniciantes. *Python* é uma linguagem de tipagem dinâmica, e é muito flexível devido à sua capacidade de usar componentes modulares projetados em outras linguagens de programação. Além disso, possui uma variedade de bibliotecas e *frameworks*, facilitando a realização de tarefas específicas e aumentando a produtividade do desenvolvedor (SRINATH, 2017).

Nesta dissertação, foi utilizada a linguagem de programação *Python* com suas bibliotecas criadas para análise de dados. Uma das bibliotecas utilizadas foi a *Pandas*. A biblioteca *Pandas* é uma das principais ferramentas para análise de dados em *Python* (PANDAS, 2020). Ela foi desenvolvida por Wes McKinney em 2008 (MCKINNEY, 2012). A biblioteca oferece estruturas de dados simples projetadas para tornar a análise mais simples, rápida e intuitiva. Ela fornece estruturas de dados de alto desempenho e fáceis de usar, permite realizar operações como filtragem, agrupamento e agregação de forma rápida e eficiente. Além disso, a biblioteca *Pandas* também fornece recursos que facilitam a análise de dados ausentes e repetidos, bem como a possibilidade de ler e gravar arquivos de dados em vários formatos, como CSV, xlsx, xls e JSON.

Outra biblioteca utilizada foi a *NumPy* (NUMPY, 2020). Ela fornece um modo eficaz de lidar com grandes volumes de dados numéricos na forma de matrizes multidimensionais. O *NumPy* possui capacidade de executar operações matemáticas de maneira vetorizada, além de fornecer funções para realizar operações de álgebra linear, como resolver sistemas de equações lineares, fatorar matrizes e calcular valores e autovetores. A biblioteca *NumPy* é uma ferramenta poderosa e essencial para quem trabalha com dados numéricos em *Python*.

*Matplotlib* é uma biblioteca para criar visualizações estáticas e interativas em *Python* (MATPLOTLIB, 2020). Possui uma integração nativa com outras bibliotecas *Python* populares, como *NumPy*, *Pandas* e *SciPy*. A biblioteca *Seaborn* também foi usada para visualização dos dados gerados (WASKOM, 2021).

A biblioteca *Spearmanr* é uma extensão da biblioteca *SciPy* (VIRTANEN *et al.*, 2020). A biblioteca *Spearmanr* fornece uma forma fácil de calcular o coeficiente de correlação de Spearman. Essa biblioteca foi usada para calcular as correlações entre as variáveis dos conjuntos de dados analisados.

O *Google Colaboratory*, ou simplesmente *Colab*, é um ambiente computacional em nuvem gratuito fornecido pela Google. O *Colab* permite aos usuários criar e compartilhar

documentos que contêm código, equações, visualizações e texto que possui um interpretador *Python*. O ambiente permite que se possa usar uma plataforma de desenvolvimento sem precisar instalar qualquer software na máquina do desenvolvedor. A alta capacidade de processamento na nuvem chama a atenção por facilitar o trabalho de análise de dados de forma rápida, prática e gratuita. O uso de [Graphics Processing Unit \(GPU\)](#) e [Tensor Processing Unit \(TPU\)](#) disponibilizado pelo ambiente pode ser usada para acelerar o processamento de dados ([COLAB, 2020](#)). Nesta pesquisa de mestrado, o *Colab* foi usado como ambiente de desenvolvimento. A sua facilidade de uso e colaboração permitiu o compartilhamento dos documentos, além de ter uma integração com o Google Drive para armazenamento e compartilhamento de arquivos.

O *Orange Canvas* é uma plataforma utilizada para aprendizado de máquina e mineração de dados ([DEMŠAR et al., 2013](#)). Desenvolvido pela Universidade de Ljubljana na Eslovênia, a plataforma possui uma interface gráfica amigável fornecida por meio de programação visual e *widgets* ([DEMŠAR et al., 2004](#)). O *Orange Canvas* possui ferramentas de mineração de dados, incluindo algoritmos de classificação, agrupamento e regressão. A ferramenta também fornece mapas, gráficos de barras, *scatter plots* e mapas de calor, que podem ser usados para explorar e compreender os dados. O *Orange Canvas* foi usado nesta pesquisa de mestrado para realizar o agrupamento hierárquico e identificar os grupos formados pelos participantes do [ENEM](#) antes e durante o período de pandemia, considerando os dados socioeconômicos e o desempenho dos participantes.

## 4.5 Preparação dos Dados

As tarefas nesta etapa incluem seleção dos dados, limpeza, formatação e integração para gerar um novo conjunto de dados como base para a pesquisa. Durante a fase de preparação dos dados, foi criada outra coluna no conjunto de dados com as pontuações médias dos participantes. Essa coluna é composta pela soma das cinco áreas do conhecimento e dividido por cinco. O conjunto de dados original possui apenas a média de cada área, por isso foi produzida a nova coluna com a média de todas as áreas juntas. Depois que a coluna foi criada, as linhas com dados vazios, as quais representam candidatos ausentes, foram removidas.

Durante a preparação dos dados foi feito o agrupamento das regiões utilizando a variável da `SG_UF_PROVA`, que contém a Sigla da Unidade da Federação da aplicação da prova. Para essa etapa, foram criados conjuntos de dados contendo os estados de cada região. Esse novo conjunto de dados foi criado para agrupar os participantes de cada estado, portanto, só possuía sete colunas, cinco para cada área do conhecimento, uma com a média geral das cinco áreas e uma que representava o estado. Esse conjunto de dados possuía 27 linhas, uma para cada estado do Brasil.

As variáveis do questionário socioeconômico com a nota média foram utilizadas para criar um novo conjunto de dados. Esse novo conjunto de dados foi feito com o intuito de analisar melhor esses dados e gerar o modelo de agrupamento e correlação. O modelo de regressão utilizou as 25 questões socioeconômicas com o intuito de encontrar as correlações com a variável de nota média criada.

Para a criação de modelos de agrupamento hierárquico, as variáveis Q001 e Q002 foram agrupadas para uma variável denominada escolaridade dos pais. As variáveis Q022, Q024, Q025 que tratam de celular, computador e internet foram agrupadas em uma única variável chamada de “acesso à tecnologia”. Para esse modelo, também foi utilizada a variável Q006, que consta com a renda familiar do participante, a qual sofreu uma alteração para o desenvolvimento dos modelos de agrupamento. Para isso, foi criada uma condicional que recebia os valores de baixa, média e alta renda. O resultado de baixa renda foram atribuídos aos participantes com renda inferior a 3 salários mínimos; o valor de média renda foram atribuídos aos participantes com renda maior que 3 salários mínimos e menor ou igual a 6; e a alta renda foi dada para os participantes com renda superior a 6 salários mínimos.

Nesta etapa de preparação dos dados, também foi utilizado o dicionário de microdados do [ENEM](#), uma ferramenta disponibilizada pelo [INEP](#) para auxiliar no processo de compreensão das colunas a serem exploradas. Vale ressaltar que o processo de preparação de dados é crucial para a fase de modelagem dos dados. Portanto, os conjuntos de dados de todos os anos passaram pelo mesmo processo de preparação.

## 4.6 Visualização e Modelagem dos Dados

A etapa de visualização e modelagem refere-se ao processo crítico de realizar investigações em dados para descobrir padrões, detectar anomalias, testar hipóteses e verificar suposições com a ajuda de estatística descritiva 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 dos dados, o principal foco dessa etapa foi gerar informações comparativas e analisar se houve impacto, e qual foi ele.

Do conjunto total de dados, foram selecionadas 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, os quais são caracterizados por estarem presente nos dois dias de prova. Após a seleção das características abordadas, geramos diversas modelagens, a saber: correlação, agrupamento hierárquico e análises estatísticas descritivas, com o uso de gráficos e tabelas, individualmente para cada um dos anos (i.e., cinco *notebooks* no *Google Colaboratory*). Após

a análise inicial de cada ano, gerou-se uma análise (i.e., um sexto *notebook*), carregando os microdados dos cinco anos para propósitos de comparação.

## 4.7 Conclusão

Este capítulo apresentou a metodologia utilizada e como cada etapa da metodologia foi usada. Este capítulo mostrou como as etapas foram abordadas durante a pesquisa, mostrando como foi feito o pré-processamento dos dados até a etapa de modelagem. A escolha do [CRISP-DM](#) ajudou a fornecer análises mais precisas devido as suas características de interatividade. Por fim, a metodologia [CRISP-DM](#) ajudou no processo de descoberta de conhecimento, o qual será apresentado no capítulo seguinte.

## 5 Resultados

Este capítulo tem o objetivo de apresentar os resultados para as questões de pesquisa. Os resultados são apresentados de forma individual para cada questão de pesquisa e mostram os resultados de maneira a comparar o período antes e durante a pandemia.

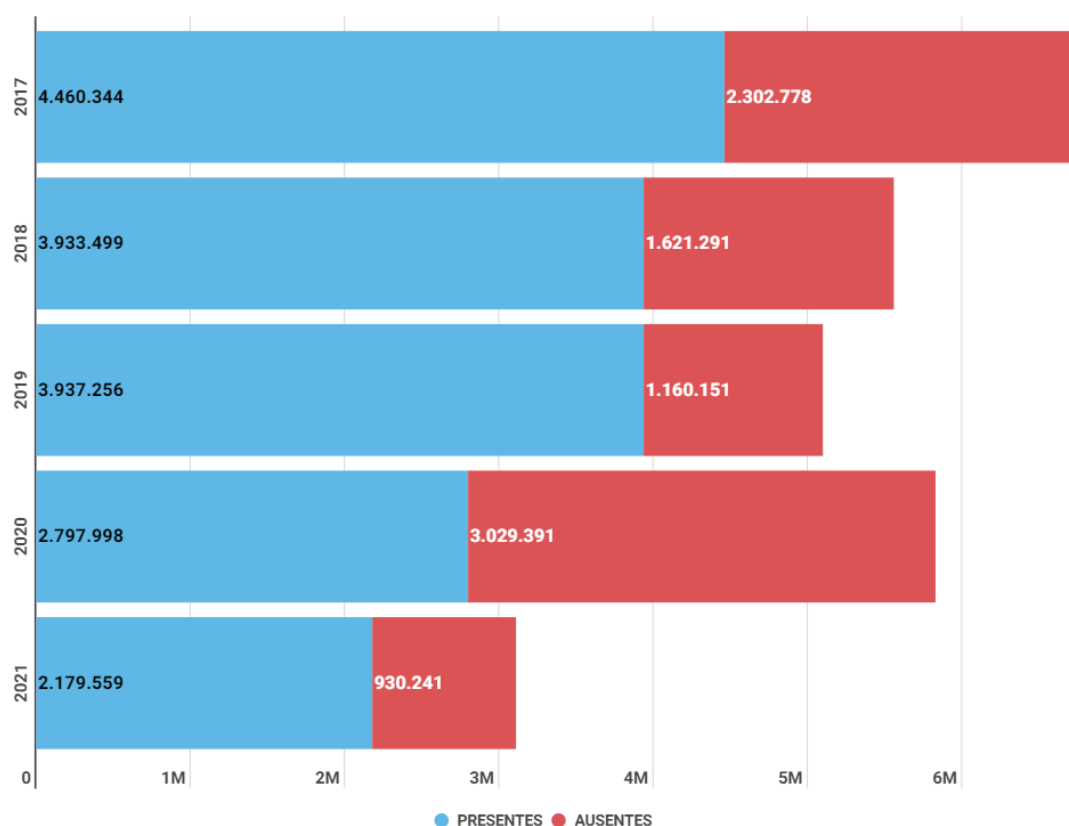
### 5.1 Impacto no Número de Participantes

A Figura 5 mostra a quantidade de ausentes (i.e., faltou pelo menos um dia) e presentes (i.e., realizou a prova os dois dias) ao longo dos 5 anos. A edição de 2017 possui a maior quantidade de candidatos presentes e a segunda maior quantidade de candidatos ausentes. Na edição de 2018, a quantidade de ausentes diminuiu em 681.487 participantes ausentes quando comparado com o ano de 2017. No ano de 2019, a quantidade de presentes teve um aumento comparado ao ano de 2018 e uma diminuição comparada ao número de ausentes. Pode-se notar que entre os anos anteriores à pandemia, os quais chamaremos de “pré-pandemia” nesta pesquisa, o exame passava por uma diminuição no número de inscritos. No ano de 2020, o ENEM, que era normalmente realizado entre os meses de outubro e novembro, foi realizado em janeiro de 2021, devido à pandemia. Os órgãos responsáveis pela realização da prova precisaram de mais tempo para se organizar, por isso, o atraso na aplicação da prova. A edição do exame de 2021, voltou a ocorrer no mês de novembro. No exame referente ao ano de 2020, que ocorreu já em 2021, ou seja, no primeiro exame ocorrido durante a pandemia, o número de inscritos totais voltou a subir, sendo maior que nos 2 anos anteriores.

No primeiro ano de pandemia, o número de inscritos foi maior do que em 2019. A edição de 2020 começou a receber inscrições em maio de 2020. No mês de maio, o Brasil se tornava o segundo país com mais casos confirmados de COVID-19, atrás apenas dos Estados Unidos. No mês de maio foram registrados 59.010 mortes pela COVID-19. Apesar de um aumento na disseminação da doença e o número de mortos aumentando exponencialmente, a edição de 2020 superou a quantidade de inscritos do ano de 2018 e 2019. Antes mesmo da abertura das inscrições em maio, o artigo (AQUINO *et al.*, 2020) mostra que a maioria dos estados já tinha determinado fechamento parcial ou total das escolas no mês de março. Isso mostra que, apesar das consequências que a doença poderiam causar, os participantes ainda se sentiram confortáveis para se inscrever. Entretanto, a Figura 5 mostra que em 2020 também foi a maior taxa de abstenção entre os anos analisados: cerca de 52% dos participantes inscritos não foram realizar a prova.

No exame de 2021, o número de inscritos foi o menor dentre os anos analisados. A

Figura 5 – Inscritos presentes, ausentes e totais para as provas do ENEM nos 5 anos analisados.



Fonte: Elaborado pelo autor (2023).

quantidade de inscritos em números absolutos foi abaixo do comum, comparado ao ano de 2009, ano da “vestibularização” do ENEM. A diferença é de 4.148.721 inscritos em 2009 para 3.109.800 em 2021. A porcentagem de abstenção diminuiu para aproximadamente 29,9%, correspondendo a uma abstenção menor que 2019. Em números absolutos, a quantidade de inscritos ausentes foi o menor de todos os anos analisados, com apenas 930.241 candidatos ausentes. Uma hipótese acerca da diminuição da abstenção no ENEM em 2021 em relação ao ano anterior é que os inscritos estavam mais confiantes e seguros para realizar o exame neste ano. É possível que isso tenha ocorrido devido a uma série de fatores, como o aumento das medidas de segurança sanitária durante a pandemia e a maior disponibilidade de informações sobre o ENEM.

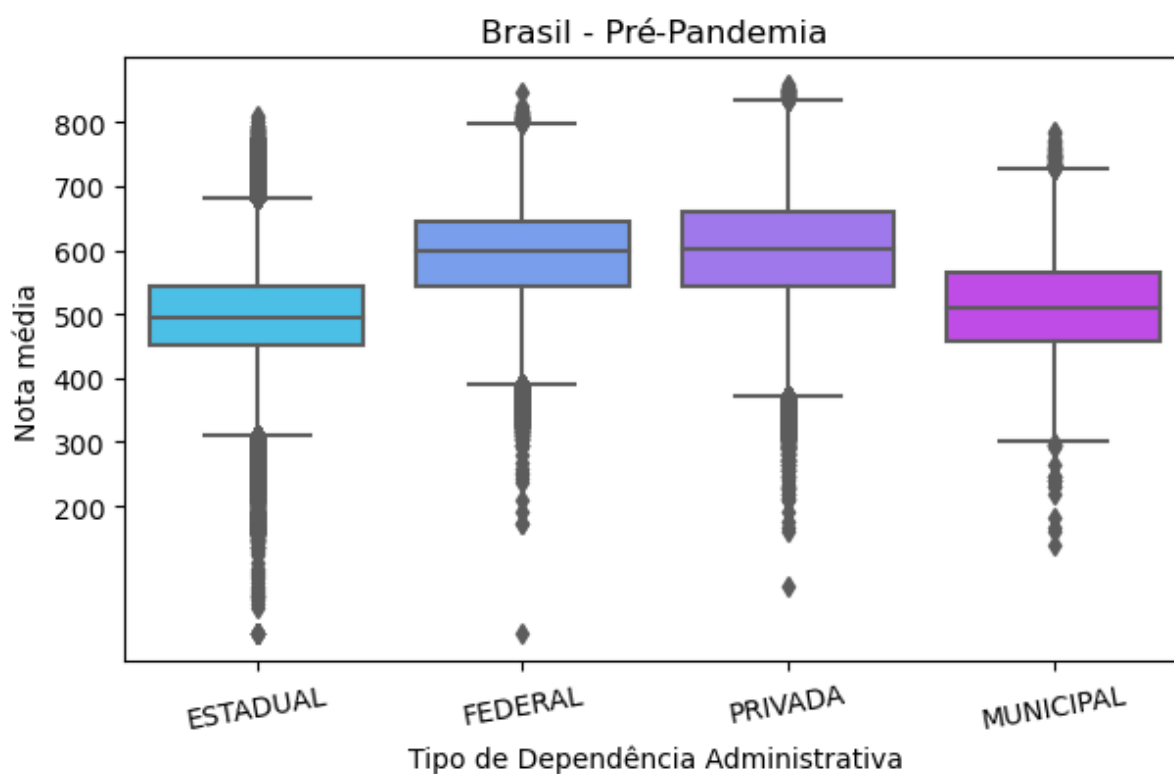
## 5.2 Impacto no Desempenho Considerando a Dependência Administrativa da Escola

Essa seção tem como intuito analisar o desempenho dos participantes do ENEM em relação ao tipo de dependência administrativa da escola do participante. Para analisar

esse caso, o conjunto de dados de 2017, 2018 e 2019 foi agrupado em um único conjunto de dados. O mesmo processo de agrupamento foi feito para as edições de 2020 e 2021. Sendo assim, a seção apresenta os dados nomeados como “pré-pandemia” e “pandemia”.

A Figura 6 apresenta em um gráfico de *boxplot* o desempenho dos participantes presentes durante as edições de 2017, 2018 e 2019 do ENEM em relação ao tipo de dependência administrativa da escola que estudam ou, em caso de participantes concluintes, que estudaram. Na Figura é possível ver os quartis e analisar o desempenho dos participantes. Podemos também analisar a diferença entre as notas das escolas particulares e das públicas. Durante a pré-pandemia, podemos notar que as escolas privadas têm desempenho superior as demais. A Figura mostra que as escolas estaduais possuem uma maior quantidade de *outliers* tanto positivos quanto negativos. As escolas municipais são as que possuem menor quantidade de *outliers* negativos e as escolas privadas, apesar de terem a mediana maior dentre as demais, possuem menor quantidade de *outliers* positivos.

Figura 6 – Desempenho dos participantes conforme o tipo de dependência administrativa durante os três anos de pré-pandemia analisados.



Fonte: Elaborado pelo autor (2023).

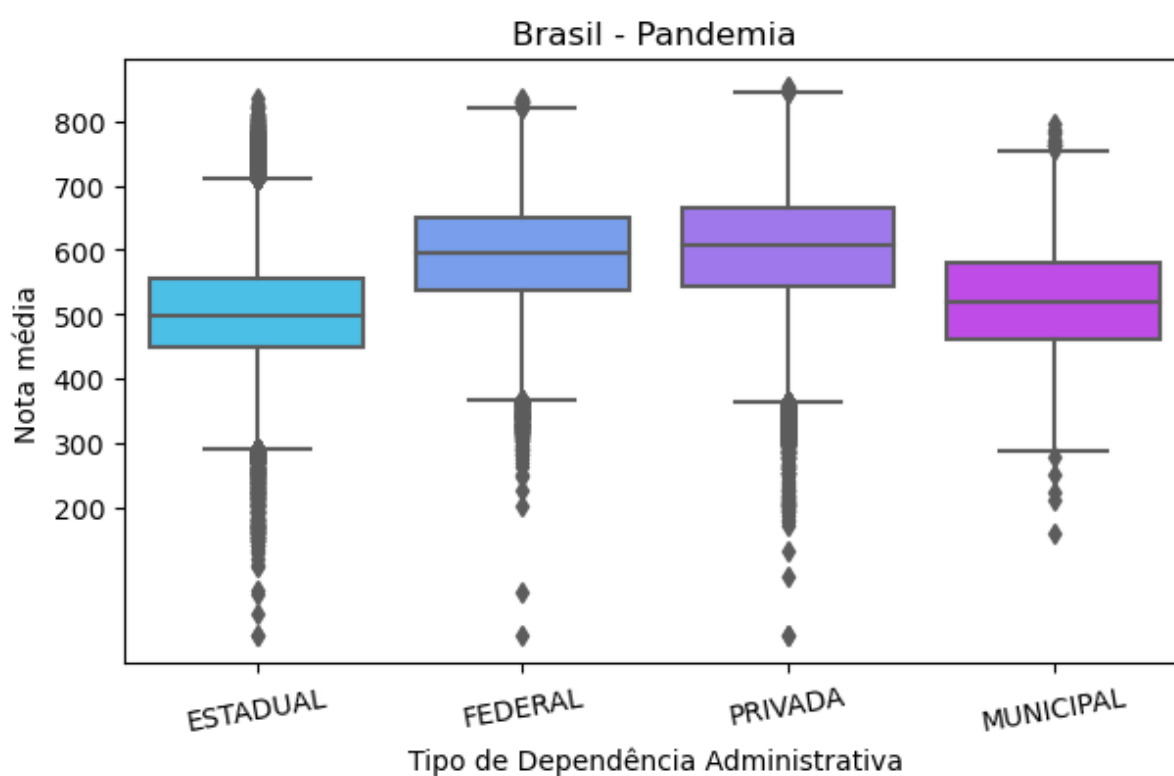
A Figura 7 mostra as notas médias em relação ao tipo de dependência administrativa durante o período da pandemia. A Figura mostra que as escolas públicas (i.e., estaduais, municipais e federais) possuem medianas menores do que as escolas privadas. É possível notar que os *outliers* positivos são maiores, independente do tipo de escola. É possível notar também que as escolas de dependência municipal são as que possuem menor quantidade



de *outliers*. Durante o período da pandemia, foi possível notar uma maior quantidade de *outliers* próximos ou igual a zero. Apesar da diminuição na quantidade de participantes e da mudança do ensino para um ambiente à distância, quando comparado o desempenho aos anos de pré-pandemia analisados, pode-se notar um leve aumento no desempenho dos participantes.

A disparidade entre as notas dos candidatos que vêm de escolas públicas e privadas é ainda visível durante a pandemia. Até a edição de 2019, o [INEP](#) fornecia um número identificador das escolas. Esse número era utilizado para avaliar as melhores escolas do Brasil em relação ao [ENEM](#). Entretanto, a partir de 2020 e em razão de atender as normas da [LGPD](#), o instituto decidiu retirar essa coluna do conjunto de dados. Em 2019, as sete primeiras escolas com maior nota média são de dependência privada. As escolas estaduais e municipais só aparecem na posição de número 166 e 607, respectivamente, e uma escola federal aparece na oitava posição.

Figura 7 – Desempenho dos participantes conforme o tipo de dependência administrativa durante a pandemia da COVID-19.



Fonte: Elaborado pelo autor (2023).

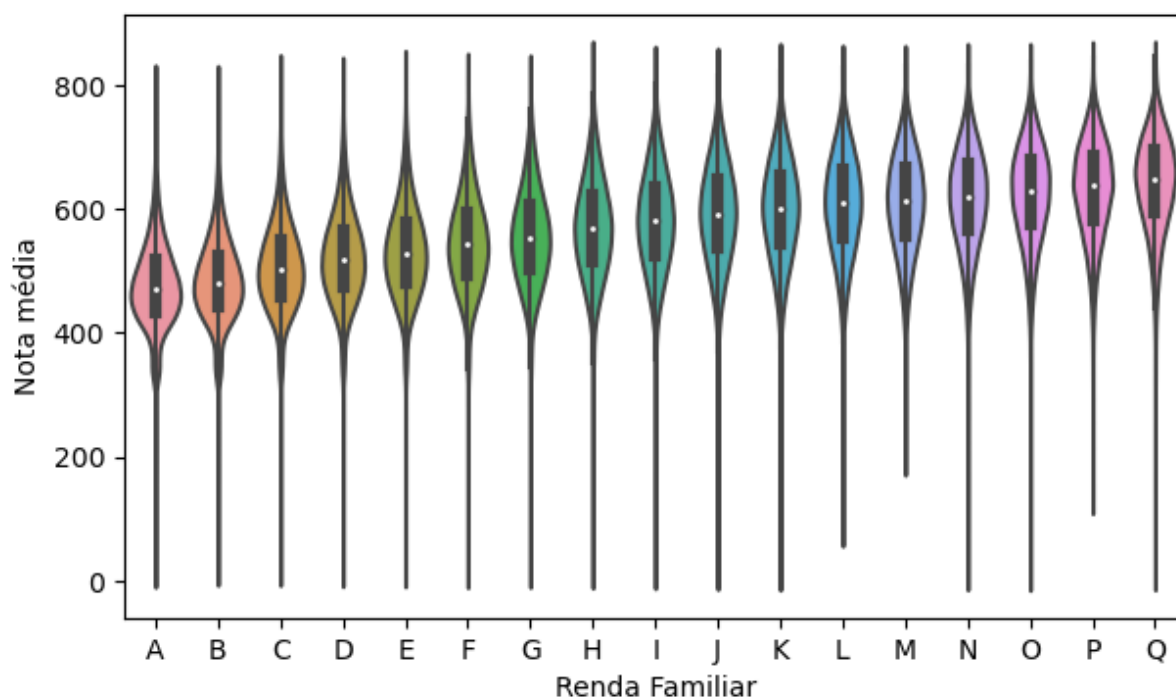
### 5.3 Impacto no Desempenho Considerando a Renda Familiar

Essa seção visa analisar os impactos da pandemia para participantes do [ENEM](#) em relação à renda familiar declarada, pois, como apresentado no capítulo 3, a renda familiar

influencia nas notas dos participantes. Assim como na questão de pesquisa apresentada anteriormente, foi feito o agrupamento dos anos de 2017, 2018 e 2019 para analisar a pré-pandemia e o agrupamento dos anos de 2020 e 2021 para apresentar os resultados da pandemia.

A Figura 8 mostra por meio de um gráfico de *violinplot* a nota média dos participantes em relação ao tipo de renda familiar dos anos de pré-pandemia. Na Figura 8, é possível notar que os valores medianos são proporcionalmente melhores a cada tipo de renda familiar. Esses valores mostram que, quanto menor é a classe social dos participantes, menor é a nota média dele. Nos anos de pré-pandemia, é possível ver que os participantes que declaram ser de renda familiar do tipo "A" (i.e., participantes com nenhum tipo de renda), estão concentrados em uma média próxima aos 500 pontos. Entretanto, vale ressaltar que existem alguns participantes dessa mesma renda que conseguiram tirar nota acima de 800 pontos.

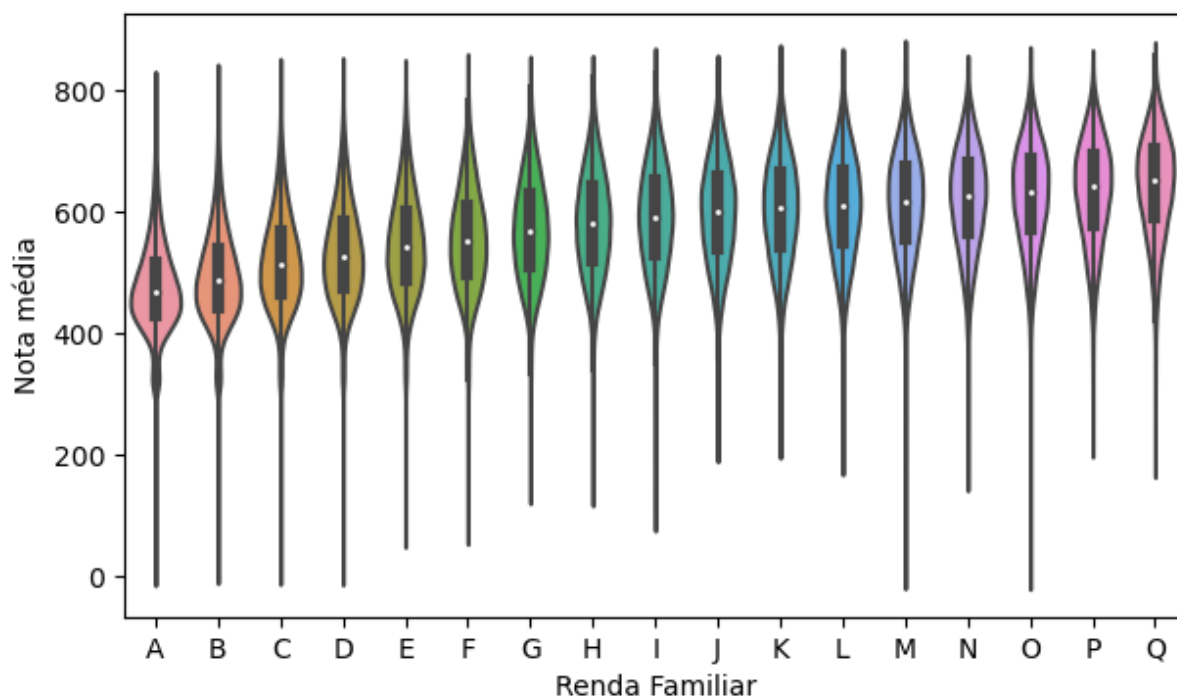
Figura 8 – Notas dos participantes do ENEM conforme a renda familiar durante o período de pré-pandemia.



Fonte: Elaborado pelo autor (2023).

A Figura 9 apresenta os dados referentes aos anos de pandemia. Apesar da aplicação da prova de 2020 já ocorrer durante o período de fechamento das escolas, houve um leve aumento na média geral desse ano. A Figura mostra que os alunos de renda "A" (i.e., participantes com nenhum tipo de renda) ainda estão concentrados em uma média próxima a 500, mas com participantes chegando em notas mais altas. É possível notar que em comparação aos anos de pré-pandemia, os resultados se assemelham, ou seja, quanto maior a renda, maior a média.

Figura 9 – Notas dos participantes do ENEM conforme a renda familiar durante o período da pandemia.



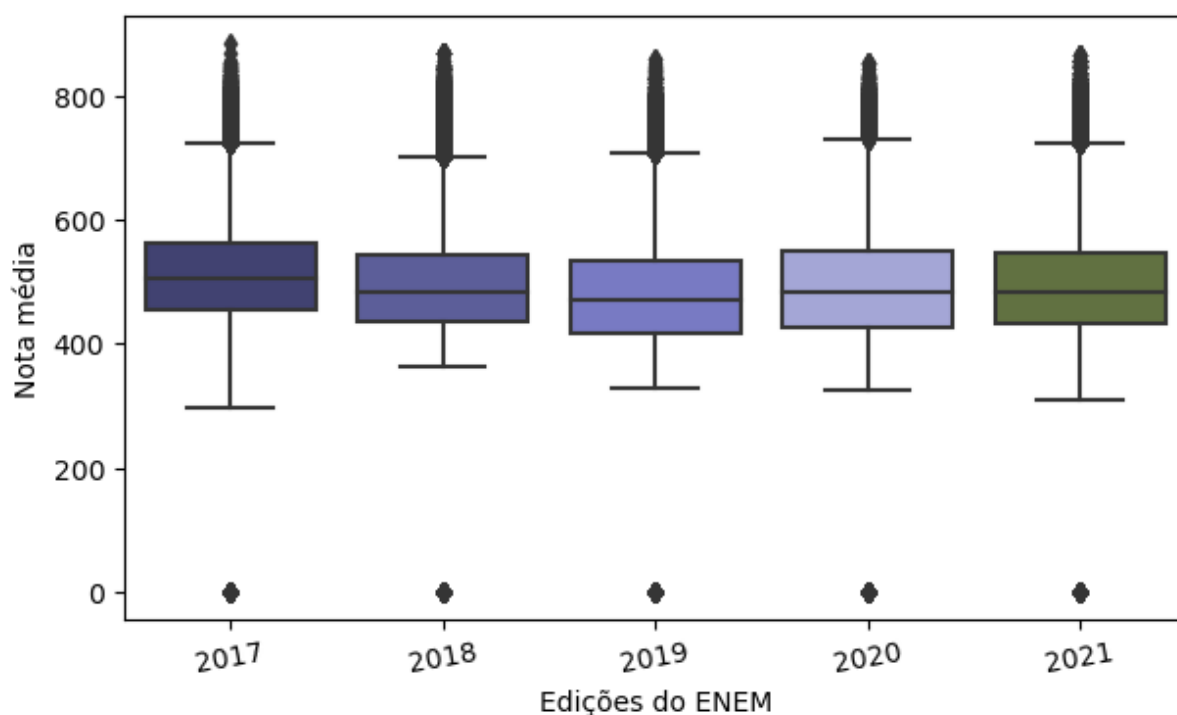
Fonte: Elaborado pelo autor (2023).

## 5.4 Áreas do Conhecimento Mais Impactadas

Para responder à questão de pesquisa: quais são as áreas do conhecimento mais impactadas durante o período da pandemia da COVID-19?, foi gerado um gráfico para cada área do conhecimento com o intuito de encontrar diferenças entre os anos e para cada área separadamente. A Figura 10 apresenta em um *boxplot* os resultados das notas da área de ciências da natureza e suas tecnologias durante os 5 anos somente com os participantes presentes nos dois dias. A maior mediana da área de ciências da natureza foi no ano de 2017, com uma mediana de 510 pontos, conforme apresentado na Figura 10. No ano de 2018 e 2019, essa nota foi baixando, alcançando 493,75 em 2018 e 477,88 em 2019. Durante a pandemia, a mediana da área voltou a subir. Em 2020, a mediana foi de 490,53, e de 493,1 em 2021.

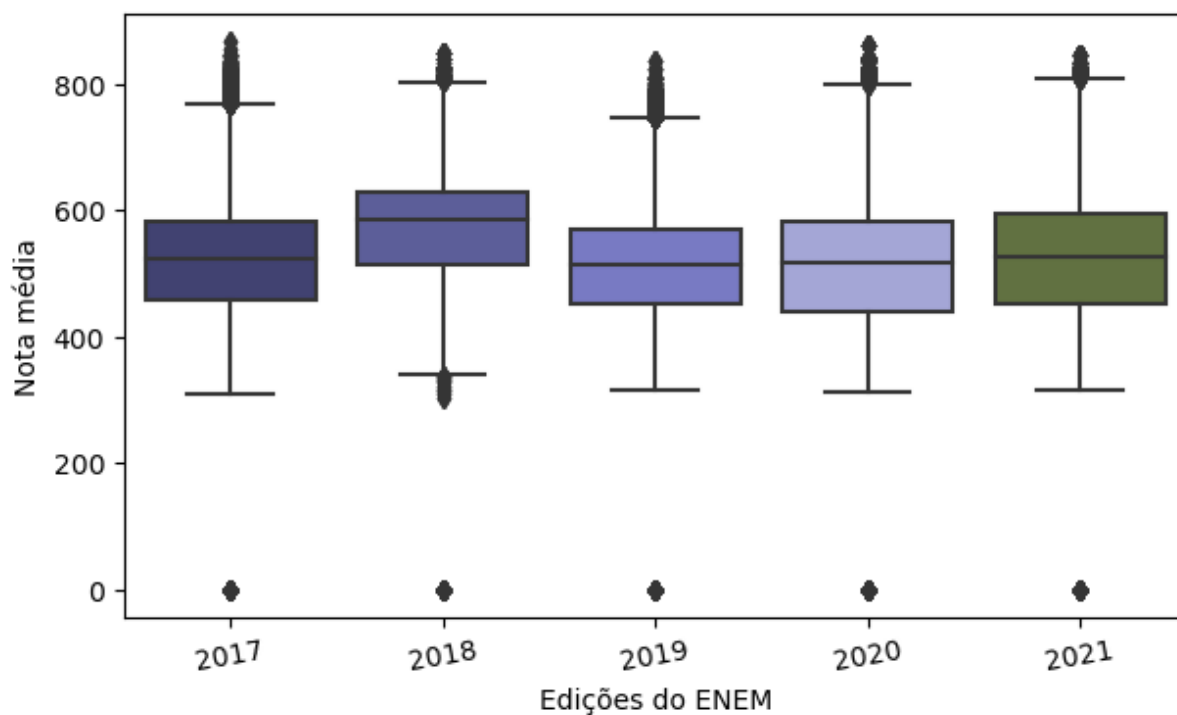
A Figura 11 apresenta os resultados da área de conhecimento de ciências humanas e suas tecnologias. Na área de conhecimento de ciências humanas e suas tecnologias, o ano de 2018 foi o que apresentou maior mediana, com 570 pontos. É possível afirmar que 2018 foi o ano com menor número de *outliers* positivos e uma maior quantidade de *outliers* negativos. O ano de 2017 teve maior quantidade de *outliers*, seguido do ano de 2019. As edições do exame durante a pandemia tiveram mediana maior que o ano de 2019. Em 2019, a mediana foi de 509,88 pontos, já em 2020 os participantes alcançaram 514,28 pontos, e 524,82 no ENEM de 2021.

Figura 10 – Notas da área de ciências naturais e suas tecnologias dos cinco anos analisados.



Fonte: Elaborado pelo autor (2023).

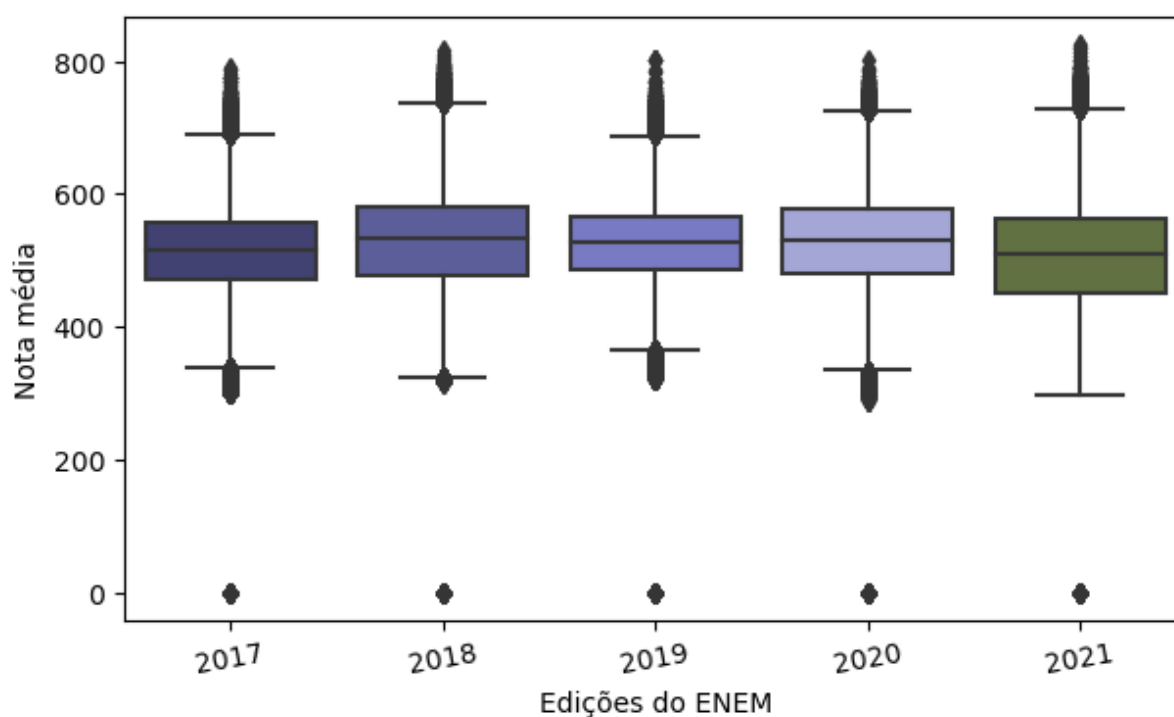
Figura 11 – Notas da área de ciências humanas e suas tecnologias dos cinco anos analisados.



Fonte: Elaborado pelo autor (2023).

As notas da área de linguagens, códigos e suas tecnologias estão representadas na Figura 12. A edição de 2021 foi o ano de menor mediana, com 506,02 pontos. Apesar desse resultado, a maior nota também foi nesse ano, com 826,1 pontos. No primeiro ano da pandemia, a mediana da área ficou em 526,01, cerca de 4 pontos a mais que no ano de 2019. Entretanto, o desvio padrão do ano de 2020 foi de 73,04, enquanto o de 2019 foi de 62,74, o que mostra que os valores amostrais de 2020 estão bem distribuídos longe da média.

Figura 12 – Notas área de Linguagens, códigos e suas tecnologias dos cinco anos analisados.

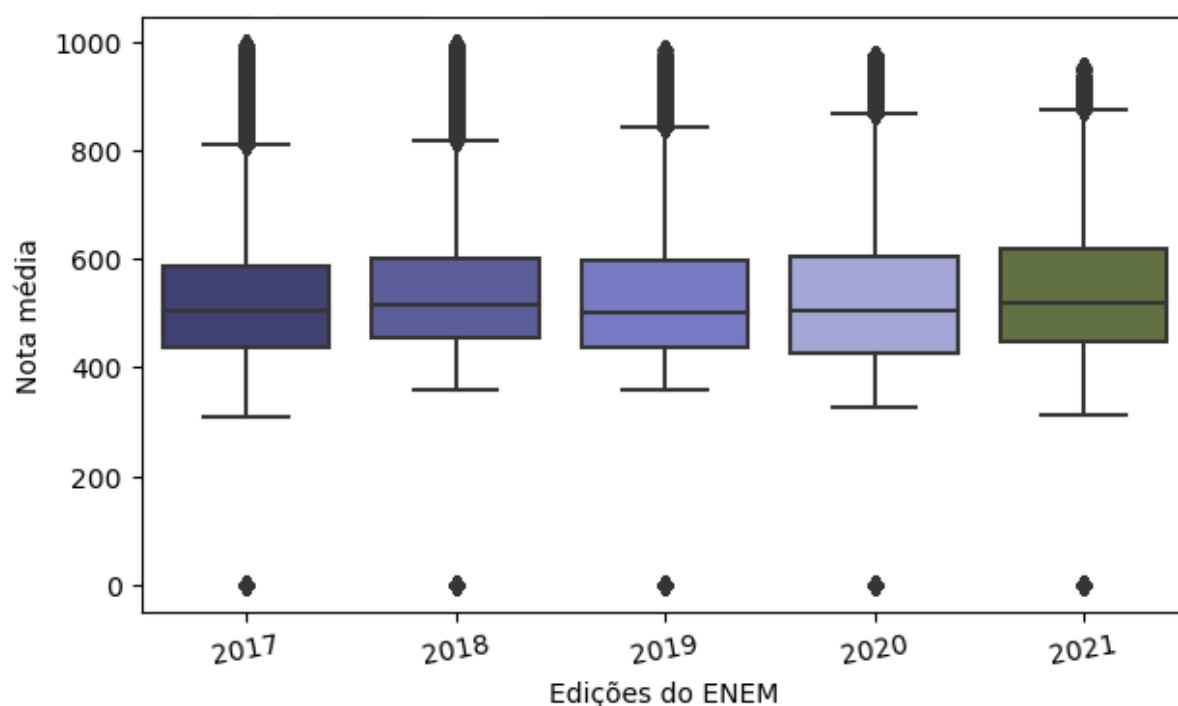


Fonte: Elaborado pelo autor (2023).

A Figura 13 mostra os resultados da área de conhecimento de matemática e suas tecnologias. Diferentemente das outras áreas do conhecimento, a matemática teve no primeiro ano de pandemia uma nota média inferior ao ano de 2019. Essa nota volta a subir no ano de 2021. A área de matemática é a que possui maior consistência entre os anos, o comportamento do desempenho se mostra quase igual em ambos os anos analisados. Na pandemia, a área teve menos *outliers* que nos anos anteriores, ou seja, isso mostra que as notas se concentraram em uma faixa mais estreita durante a pandemia, com menor variação entre os resultados.

A Figura 14 mostra os resultados da prova de redação, positivos e negativos. Os resultados mostraram que o ano de 2017 foi o que mais teve *outliers*. Na pandemia, a redação, assim como a área de matemática e suas tecnologias, teve uma baixa no desempenho médio em 2020 quando comparado a 2019. A redação possui maior área acima

Figura 13 – Notas da área de Matemática e suas tecnologias dos cinco anos analisados.



Fonte: Elaborado pelo autor (2023).

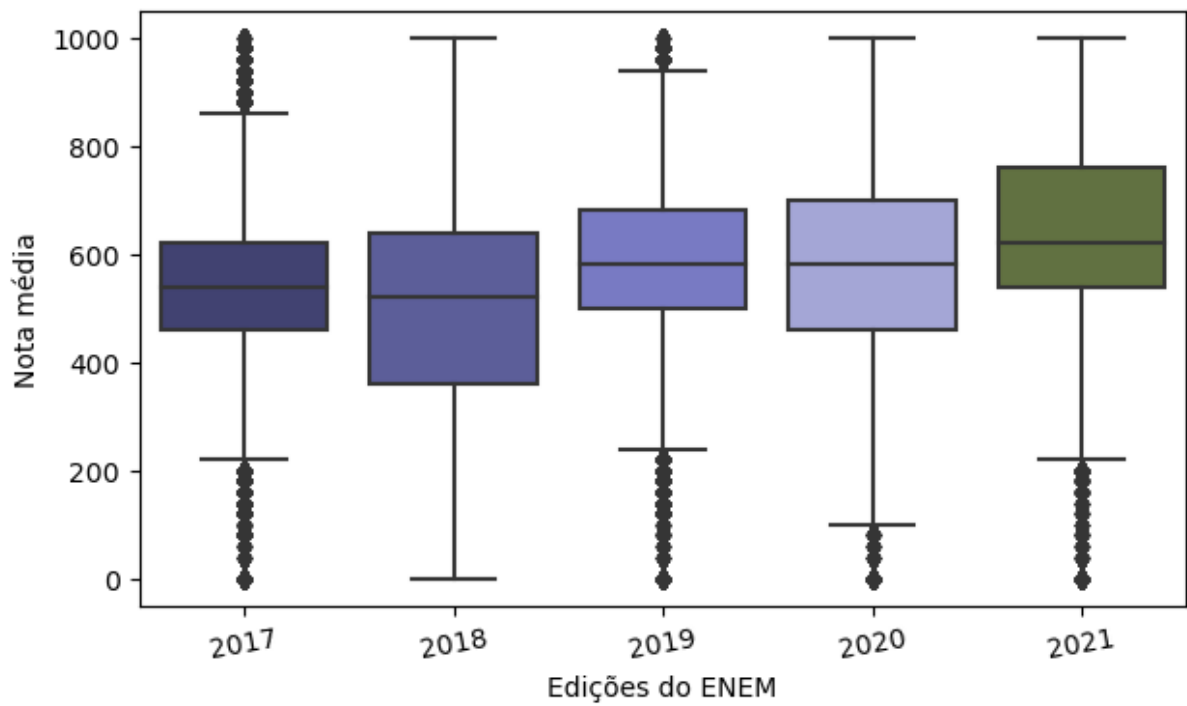
de 800. Apesar de, no primeiro ano da pandemia, a média da redação ter diminuído, no ano de 2021 a nota voltou a subir e superou os anos anteriores analisados.

## 5.5 Correlação das Características Socioeconômicas com o Desempenho

Essa seção tem o objetivo de apresentar os resultados para a QP5. Utilizando um modelo de correlação, essa questão visa descobrir quais características socioeconômicas tem maior influência na nota dos participantes do ENEM. Para isso, foi utilizado a correlação de Spearman. A correlação de Spearman indica o grau de relacionamento entre duas variáveis, podendo conter valores entre -1 e 1. A correlação foi escolhida devido aos valores do questionário socioeconômicos poderem ser numéricos, com a não garantia de haver relação de linearidade, o que é requirido na correlação de Pearson, ou categóricos ordinais.

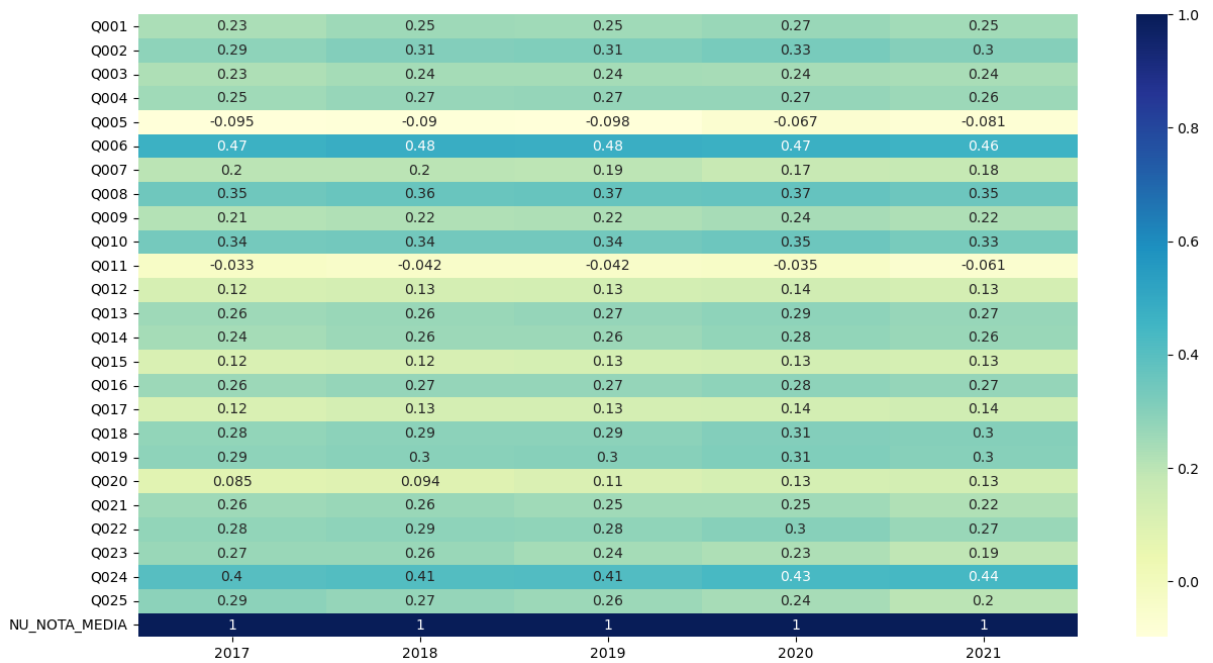
Os resultados da correlação apresentados na Figura 15 apresentam que, nos cinco anos analisados, a renda familiar é sempre a variável com maior correlação com a nota média dos participantes. Pode-se notar que as maiores correlações estão ligadas a itens de bens financeiros, como carro e aparelhos eletrônicos. Apenas duas variáveis mostraram correlações negativas, foram elas: quantas pessoas moram na sua residência? e na sua residência tem motocicleta?

Figura 14 – Notas da Redação dos cinco anos analisados.



Fonte: Elaborado pelo autor (2023).

Figura 15 – Correlação das variáveis socioeconômicas com as notas médias dos cinco anos analisados.



## 5.6 Impacto no Desempenho das Regiões e Estados

Esta seção apresenta os resultados do desempenho dos participantes de acordo com seu estado e região. Nesta seção serão apresentados o desempenho de cada estado e região e também será dado ênfase aos estados do Ceará, Maranhão e Piauí devido à sinergia entre eles.

A Figura 16 apresenta um mapa do Brasil com intervalos de valores em cada ano apresentados pela cor vermelha, onde a intensidade da cor mais forte, apresenta maior desempenho. A Figura 16 mostra que os estados da região sudeste se saíram melhor em ambos os anos analisados quando comparado com os estados da região norte. Os estados do Maranhão, Acre e Amapá tiveram notas menores a 490 pontos no ano de 2017, sendo os 3 piores estados daquele ano. Comparando as notas dos estados do Maranhão, Acre e Amapá com a região sudeste e sul, nenhum estado dessas regiões ficou com média geral menor que 510 pontos. Em 2018, as médias gerais aumentaram e os estados também tiveram aumento, mas os estados do norte e nordeste já citados mantiveram suas notas abaixo das notas da região sul, sudeste e centro oeste.

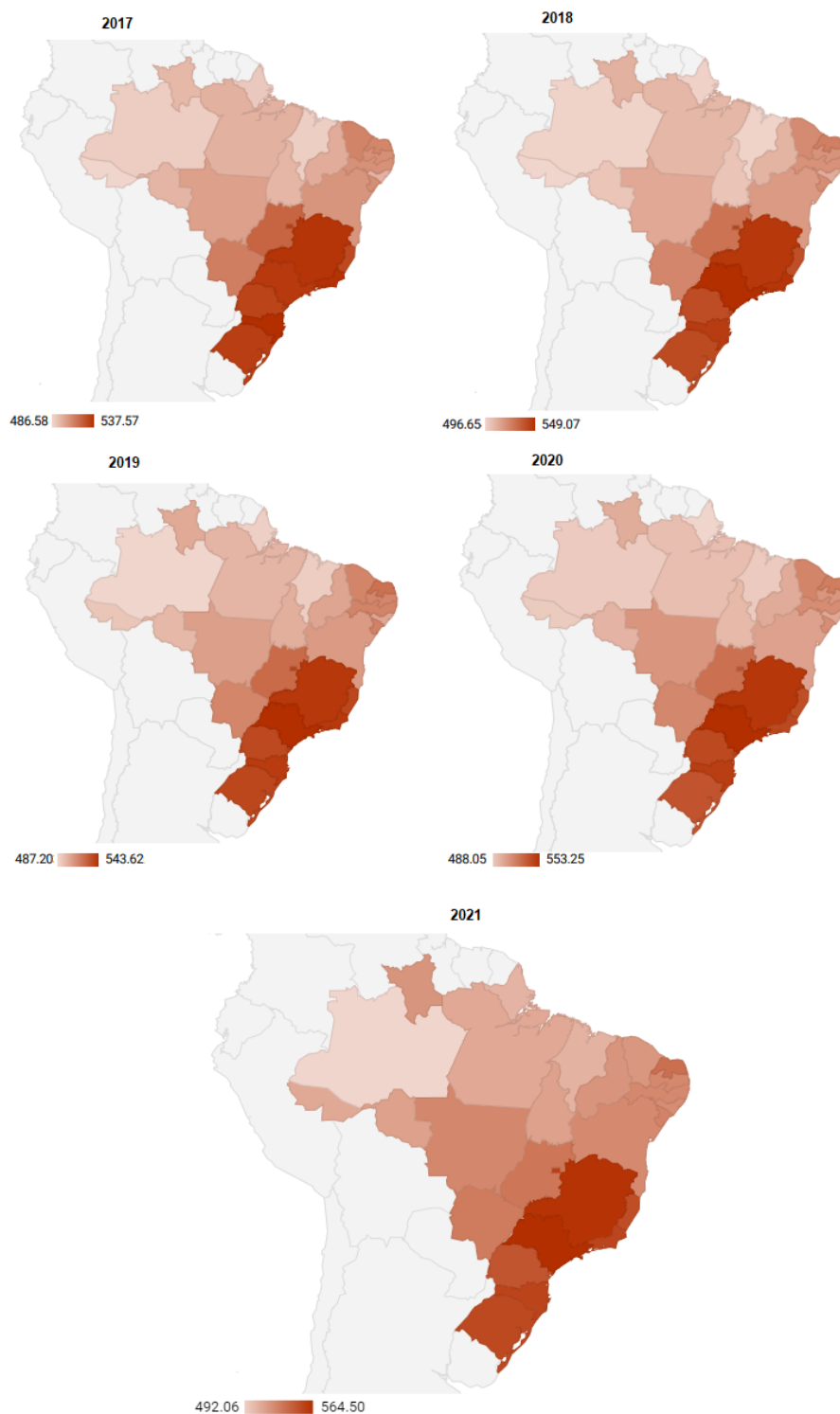
O estado com a menor nota em 2019 foi Amazonas, com média de 487,2, e São Paulo foi o que possuiu maior nota, com média de 543,62 pontos. Já em 2020 a menor média dentre os estados aumentou, indo a 488,05, sendo essa nota pertencente ao Amapá. A maior nota de 2020 também foi do estado de São Paulo, com 553,25. No ENEM de 2021, o estado de Amazonas ficou novamente com a menor média, 492,06 pontos, e o estado de São Paulo manteve novamente a maior pontuação com média de 564,5 pontos. Em 2020, os estados do Acre, Amapá, Pará, Paraíba e Tocantins tiveram desempenho abaixo de 2019. Em 2021, apenas os estados do Amazonas, Ceará e Espírito Santo tiveram médias menores que em 2020.

Durante a pandemia, a média geral aumentou gradativamente, em 2020, teve um leve aumento comparado a 2019, e em 2021 também teve um leve aumento comparado a 2020. Vale ressaltar que no ano de 2019, ainda era possível avaliar o desempenho individual das escolas, pois existia um código de identificação para cada escola. Em 2019, as duas primeiras escolas com maior pontuação do ENEM eram do estado do Ceará, sendo que as duas eram de dependência administrativa privada. Nessa classificação de melhores escolas do ano de 2019, o Piauí aparece na 11<sup>a</sup> colocação, também com uma escola privada, e o Maranhão, aparece apenas na 102<sup>a</sup> posição.

A Figura 17 apresenta, mediante uma série temporal, os resultados dos estados do Ceará, Maranhão e Piauí. Pode-se notar que, durante o período de pandemia, as notas desses 3 estados aumentaram, quando comparado aos anos anteriores. O estado do Piauí em 2021 chegou a ter uma média superior ao estado do Ceará. O estado do Maranhão também aumentou seu desempenho no período pandêmico, entretanto, ele ainda está no



Figura 16 – Mapas do Brasil com a nota média de cada estado nos 5 anos analisados.

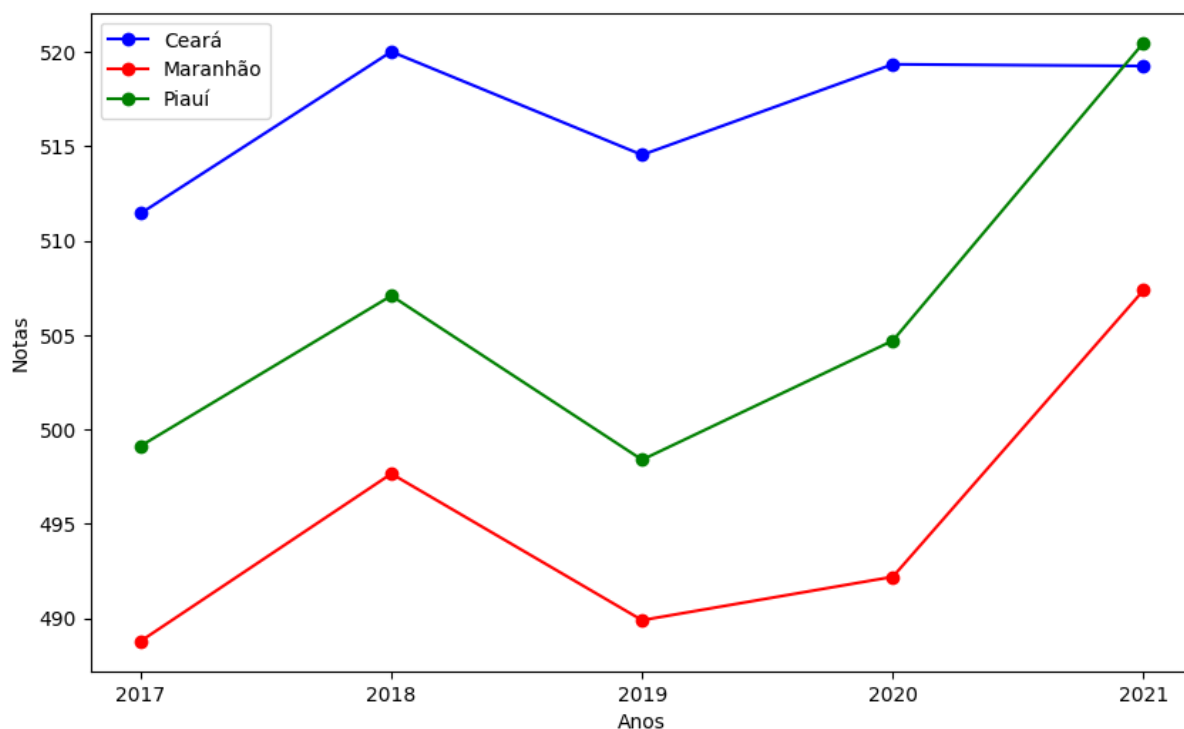


Fonte: Elaborado pelo autor (2023).

*ranking* dos estados com menor desempenho no exame.

A Tabela 5 mostra a nota de cada área do conhecimento conforme a região do Brasil na edição do ENEM de 2017. Pode-se notar que as melhores notas médias são na região sul e sudeste, e a pior média ficou na região norte. Ao analisar a Tabela 5, pode-se

Figura 17 – Série temporal do desempenho médio dos estados do Ceará, Maranhão e Piauí durante os 5 anos analisados.



Fonte: Elaborado pelo autor (2023).

notar que a região norte teve um pior desempenho em todas as áreas do conhecimento. Na região sul e sudeste, a área de conhecimento com a menor média foi a de ciências naturais e suas tecnologias, com 524,7 e 522,34 respectivamente. Já na região norte, a menor nota ficou com a área de matemática e suas tecnologias. Além disso, a média da redação é maior em todas as regiões, exceto na região norte, onde a maior média foi da área de ciências humanas e suas tecnologias.

Região	Linguagens	Matemática	Redação	Humanas	Naturais	Média
Norte	491,5	486,9	494,56	498,9	493,6	493,1
Nordeste	497,6	501,13	517,3	505,68	499,14	504,19
Centro-oeste	506,79	511,82	517,74	515,8	510,05	512,44
Sul	527,02	539,2	540,4	538	524,7	533,8
Sudeste	524,96	538,69	547,45	536,6	522,34	534,01

Tabela 5 – Desempenho geral dos participantes por região no ENEM de 2017.

A Tabela 6 apresenta os resultados de cada região do Brasil em relação às áreas do conhecimento no ano de 2018. Assim como no ano de 2017, a região norte teve a menor nota média. As notas de todas as regiões tiveram aumento quando comparados ao ano de 2017. Vale ressaltar que, do ano de 2017 para o ano de 2018, houve um aumento de

aproximadamente 10 pontos na média geral do Brasil no [ENEM](#).

Região	Linguagens	Matemática	Redação	Humanas	Naturais	Média
Norte	506,6	505,4	473,3	547,5	474,1	501,4
Nordeste	513,1	519,8	501,5	555,8	480,8	514,2
Centro-oeste	527,7	530,2	507	568,6	494,3	525,5
Sul	543	548,7	524,2	587,3	505	541,6
Sudeste	542,2	554,1	538,6	585,2	506,4	545,3

Tabela 6 – Desempenho geral dos participantes por região no [ENEM](#) de 2018.

No ano de 2019, a média geral dos participantes do [ENEM](#) diminuiu cerca de 6 pontos. Esses valores são refletidos na média das regiões conforme a Tabela 7. Podemos notar que a média da região norte, a qual já era a menor entre as regiões, teve uma redução de aproximadamente 6 pontos. A região centro-oeste teve uma diminuição de 11 pontos quando comparado ao ano de 2019. A região sul foi a que menos teve queda de desempenho no ano de 2019, tendo uma diminuição da média geral em apenas 1,58 pontos.

Região	Linguagens	Matemática	Redação	Humanas	Naturais	Média
Norte	503,09	487,31	541,12	487,83	456,46	495,16
Nordeste	509,21	506,01	571,55	494,43	464,89	509,22
Centro-oeste	517,12	514,2	564,38	503,19	473,28	514,44
Sul	535,65	545,01	601,32	525,482	492,62	540,02
Sudeste	536,06	543,22	585,21	527,55	490,66	536,54

Tabela 7 – Desempenho geral dos participantes por região no [ENEM](#) de 2019.

No ano de 2020, todas as regiões tiveram aumento na sua nota média, conforme a Tabela 8. Isso mostra que, no primeiro ano de pandemia, não houve impacto negativo para as regiões, quando se analisando somente a média geral. As regiões norte e nordeste tiveram um aumento de apenas 1,05 pontos e 1,6 pontos, respectivamente. A região sudeste foi a que teve maior aumento na nota média. Os 536,54 pontos do ano de 2019 aumentaram em um pouco mais de 10 pontos em 2020, ficando com 546,88. No ano de 2020, a região sudeste foi a que obteve maior nota média. A região norte continuou com menor média.

A edição do [ENEM](#) de 2021 também teve um impacto positivo em relação às notas quando comparadas aos anos de pré-pandemia, e também melhores notas quando comparado ao primeiro ano da pandemia. A região sudeste obteve o melhor desempenho entre as regiões, com 558,1 pontos aproximadamente, como mostra a Tabela 9. Seguido da região sudeste, a região sul e centro oeste mantiveram suas classificações quando comparados

Região	Linguagens	Matemática	Redação	Humanas	Naturais	Média
Norte	501,81	481,97	538,94	488,87	469,46	496,21
Nordeste	510,32	499,28	572,05	495,65	476,79	510,82
Centro-oeste	522,4	512,31	567,13	509,19	487,48	519,7
Sul	541,92	545,27	587,04	535,27	505,04	542,91
Sudeste	542,8	546,73	604,07	534,02	506,79	546,88

Tabela 8 – Desempenho geral dos participantes por região no ENEM de 2020.

aos anos anteriores analisados. As regiões norte e nordeste, apesar de aumentarem suas notas em 2021, assim como em 2020, ainda obtiveram notas menores que as demais regiões.

Região	Linguagens	Matemática	Redação	Humanas	Naturais	Média
Norte	482,05	501,3	594,7	500,49	472,2	510,18
Nordeste	488,9	518,54	623,8	506,8	478,9	523,4
Centro-oeste	498,8	528,8	619,2	517,07	487,6	530,3
Sul	520,29	556,0	635,1	541,5	507,3	552,06
Sudeste	523,1	561,35	654,3	543,2	508,77	558,1

Tabela 9 – Desempenho geral dos participantes por região no ENEM de 2021.

Conforme apresentado na Seção 5.4, algumas áreas do conhecimento possuem *outliers* positivos. Dessa forma, é comum encontrar participantes de estados e regiões que têm médias gerais mais baixas atingirem notas altas em algumas das áreas do conhecimento. Por exemplo, em 2021 a região nordeste é a recordista em nota 1000 na redação. Das 10 notas 1000 tiradas em 2021, sete eram da região nordeste.

## 5.7 Impactos nas Características dos Participantes

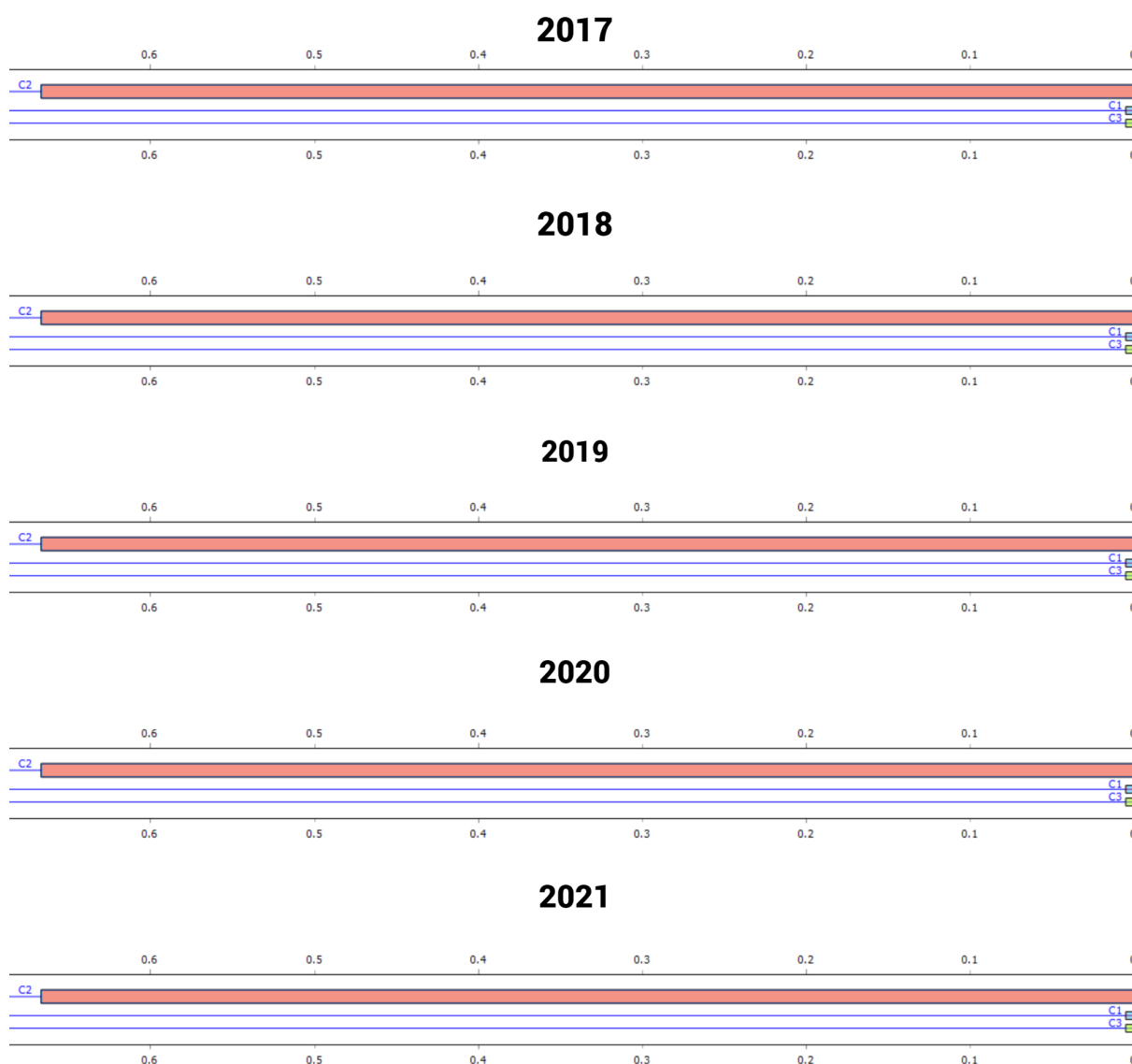
Essa seção tem o intuito de responder à sétima questão de pesquisa. A questão visa entender se as características dos participantes do ENEM mudaram durante o período de pandemia. Para isso, desenvolvemos modelos de agrupamento hierárquico, utilizando a distância de *Hamming* para medir a dissimilaridade. Foi utilizado o *linkage* completo para a distância entre os clusters, pois é menos sensível a *outliers* e ruídos. Para criar o agrupamento, foi utilizado uma amostra aleatória de 10.000 participantes para cada ano. A escolha da amostragem para essa questão de pesquisa é devido ao algoritmo utilizar muito processamento computacional e não é possível utilizar o conjunto de dados completos.

As variáveis utilizadas foram a renda familiar, acesso à tecnologia e escolaridade dos pais. As variáveis “acesso à tecnologia” e “escolaridade dos pais” foram criadas a partir de

outras variáveis, conforme apresentado no Capítulo 4. Após a criação dos Grupos, definimos a nomenclatura “Menos favorecidos”, “Parcialmente favorecidos” e “Mais favorecidos”. Sendo assim, os 3 tipos de grupos formados podem ser analisados de forma comparativa em busca de encontrar diferença nas características dos participantes antes e durante o período da pandemia da COVID-19. Vale ressaltar que os microdados de todos os anos aqui apresentados passaram pelo mesmo processo de transformação de dados.

A Figura 18 apresenta dendrogramas dos 5 anos analisados. O agrupamento hierárquico utilizando a amostragem de 10.000 participantes para cada ano gerou três grupos diferentes. Cada grupo contém uma quantidade de participantes. Pode-se notar que o grupo 2 (C2) é o maior em todos os anos, e os grupos 1 e 3 (C1 e C3) são os menores.

Figura 18 – Dendrogramas dos agrupamentos hierárquicos dos cinco anos analisados.



Fonte: Elaborado pelo autor (2023).

A Tabela 10 apresenta os resultados do algoritmo de agrupamento hierárquico nos dados da edição do ENEM de 2017. No primeiro ano analisado, pode-se perceber que o grupo menos favorecido e o mais favorecido possuem quase a mesma quantidade de amostras. O grupo 2 possui cerca de 44% dos participantes com renda média ou alta e 45% dos responsáveis com ensino superior. Além disso, de todos os grupos criados, o grupo 2 é o único que possui 100% dos participantes com algum tipo de acesso à tecnologia, seja ela, computador, celular ou internet na residência.

Menos favorecidos	Parcialmente favorecidos	Mais favorecidos
Grupo 1 (10,58%)	Grupo 2 (78,9%)	Grupo 3 (10,52%)
95,93% tem renda baixa	55,56% tem renda baixa	80,6% tem renda alta ou média
99% não tem nenhum responsável com ensino superior	65% não tem nenhum responsável com ensino superior	98,5% tem o pai ou a mãe com ensino superior
98% não tem acesso a nenhum tipo de tecnologia.	100% tem acesso a algum tipo de tecnologia	97,4% tem acesso a algum tipo de tecnologia

Tabela 10 – Agrupamento hierárquico do ENEM de 2017.

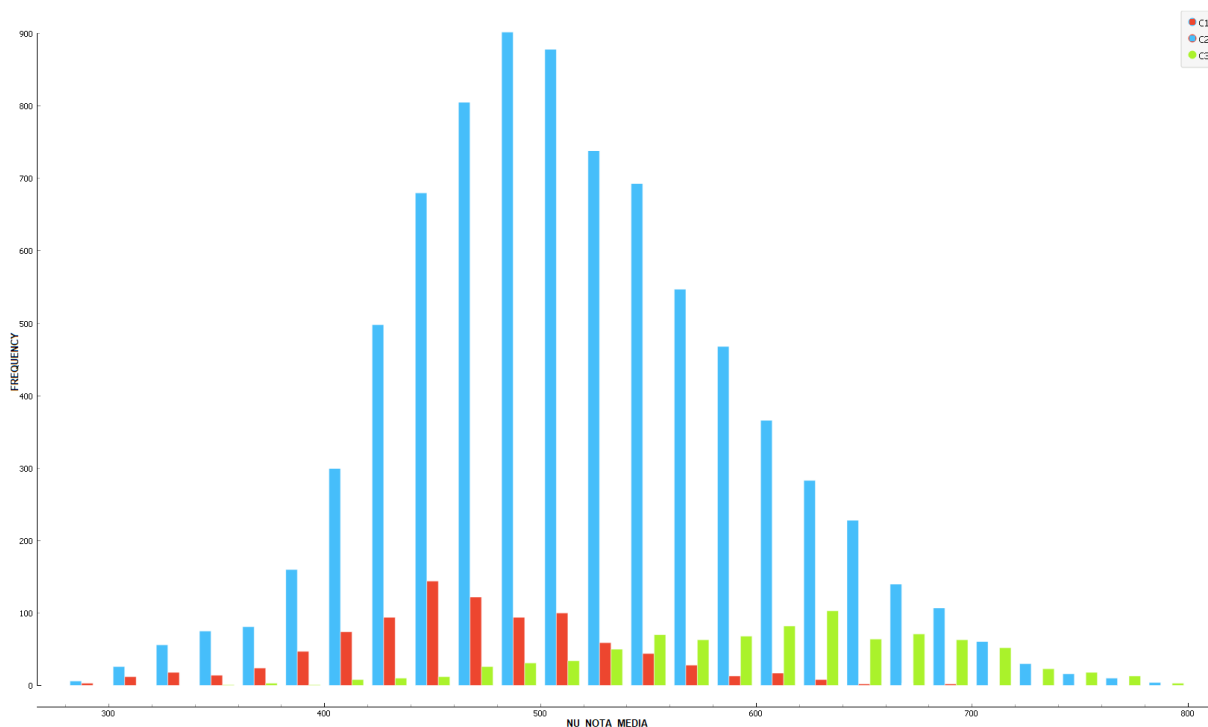
A Figura 19 apresenta a distribuição dos participantes de cada grupo em relação à nota média. É possível notar que a nota dos participantes do grupo 1 são mais baixas, enquanto as notas dos participantes do grupo 2 são distribuídas, e é possível notar a existência de mais participantes com notas mais altas em relação à concentração do grupo. Os participantes do grupo 3 são aqueles que possuem maior nota média dentre os grupos criados.

Os resultados do agrupamento no ano de 2018 estão apresentados na Tabela 11. Podemos notar que, assim como em 2017, o grupo 2 é o que possui maior quantidade de participantes. No ano de 2018, o grupo 1 foi maior que o grupo 3, repetindo o comportamento de 2017. É possível notar que o grupo 2 teve um aumento na quantidade de participantes de renda familiar baixa. Já o grupo 3, aumentou em 8% a quantidade de participantes com renda média ou alta.

A frequência dos participantes de cada grupo é para o ano de 2018 apresentada na Figura 20. Nota-se que o comportamento do gráfico é parecido ao ano de 2017. Existem mudanças na quantidade de participantes em cada média apresentada, entretanto, os grupos 3 e 2 têm os melhores desempenhos. Já o grupo 1, ou seja, o grupo menos favorecido, é aquele que apresenta um menor desempenho, tendo alguns participantes com notas acima de 600 pontos.

A Tabela 12 mostra os resultados do agrupamento no ano de 2019. A Tabela mostra que, no ano da pré-pandemia, o grupo 1 possui 9,21% do total da amostra. O grupo 2

Figura 19 – Frequência dos participantes de cada grupo em relação à nota média do ENEM no ano de 2017.



Fonte: Elaborado pelo autor (2023).

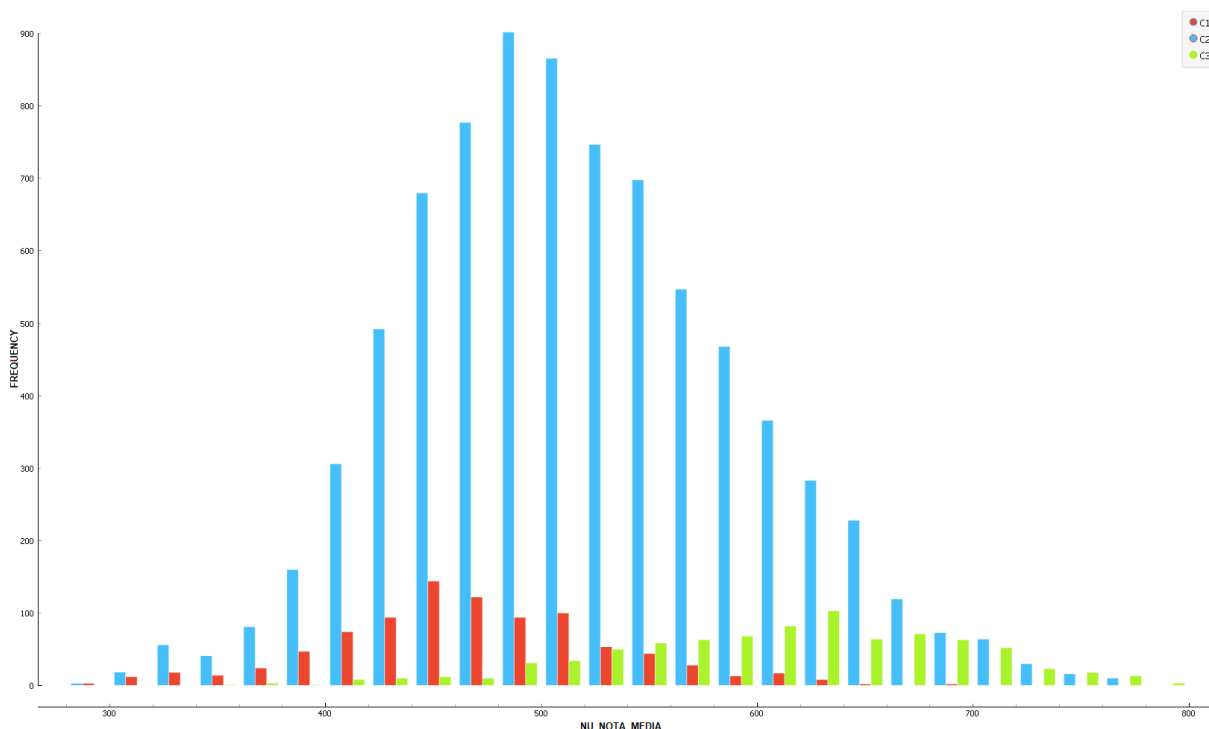
Menos favorecidos	Parcialmente favorecidos	Mais favorecidos
Grupo 1 (9,69%)	Grupo 2 (81,74%)	Grupo 3 (8,57%)
97,44% tem renda baixa	61,2% tem renda baixa	88,2% tem renda alta ou média
95,92% não tem nenhum responsável com ensino superior	74% não tem nenhum responsável com ensino superior	99,2% tem o pai ou a mãe com ensino superior
100% não tem acesso a nenhum tipo de tecnologia.	98% tem acesso a algum tipo de tecnologia	100% tem acesso a algum tipo de tecnologia

Tabela 11 – Agrupamento hierárquico do ENEM de 2018.

possui 82,1% do valor da amostra e o grupo 3 possui 8,09% da amostra. O grupo 1 é o grupo menos favorecido, ou seja, a maioria do grupo é declarado de renda baixa, 98% dos participantes não possui responsáveis com ensino superior completo e 100% não tem telefone, computador ou internet. É possível notar que as características dos grupos e a quantidade de participantes em cada um se mantêm parecido nos anos de 2017, 2018 e 2019.

A Figura 21 apresenta a distribuição de cada grupo conforme a nota média dos participantes. As barras azuis representam os participantes do grupo 2, o qual tem maior

Figura 20 – Frequência dos participantes de cada grupo em relação à nota média do ENEM no ano de 2018.



Fonte: Elaborado pelo autor (2023).

Menos favorecidos	Parcialmente favorecidos	Mais favorecidos
Grupo 1 (9,21%)	Grupo 2 (82,1%)	Grupo 3 (8,69%)
96,96% tem renda baixa	64,13% tem renda baixa	81% tem renda alta ou média
98% não tem nenhum responsável com ensino superior	80% não tem nenhum responsável com ensino superior	100% tem o pai ou a mãe com ensino superior
100% não tem acesso a nenhum tipo de tecnologia.	100% tem acesso a algum tipo de tecnologia	99% tem acesso a algum tipo de tecnologia

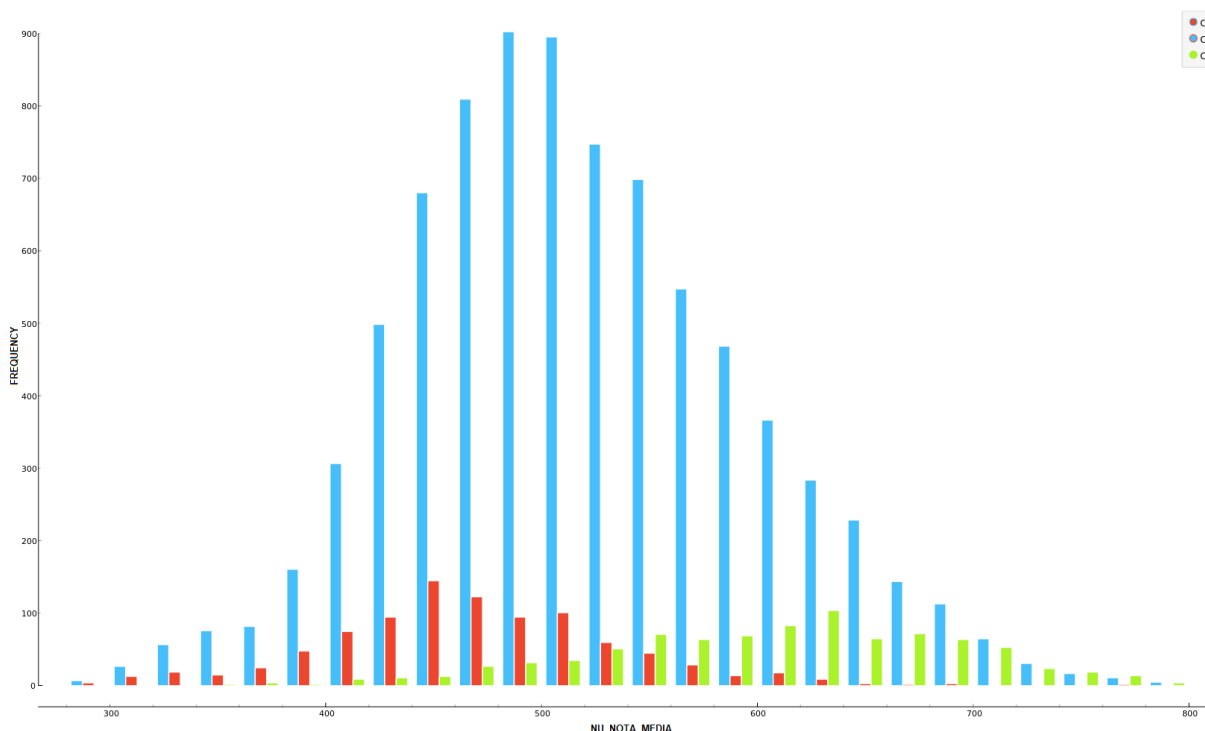
Tabela 12 – Agrupamento hierárquico do ENEM de 2019.

quantidade de participantes. As barras vermelhas mostram os participantes do grupo 1 e as barras verdes mostram os participantes do grupo 3. Pode-se notar que no ano de 2019, o grupo 1 possui menores notas médias que o grupo 2, ficando com a maior frequência em torno de 400 e 500 pontos de nota média. Já o grupo 3 possui notas médias maiores que o grupo 2. O grupo 2 fica com maior frequência em torno de 500 e 600. É possível notar que o grupo 3 tem participantes com nota próxima a 800.

A Tabela 13 apresenta os resultados do agrupamento na edição do ENEM de 2020. O grupo 1, referente aos participantes menos favorecidos, possui 7,97% do total da



Figura 21 – Frequência dos participantes de cada grupo em relação à nota média do ENEM no ano de 2019.



Fonte: Elaborado pelo autor (2023).

amostra, número inferior a 2019. Vale ressaltar que os participantes do grupo 1 também têm as mesmas características que o grupo 1 de 2019. O grupo 2 possui 64% do valor da amostra, cerca de 18% menor que no ano de 2019 e com menor quantidade dentre os 3 anos analisados anteriormente. Já o grupo 3 possui 28,03% da amostra, cerca de 19% a mais que em 2019, e maior que 2017 e 2018.

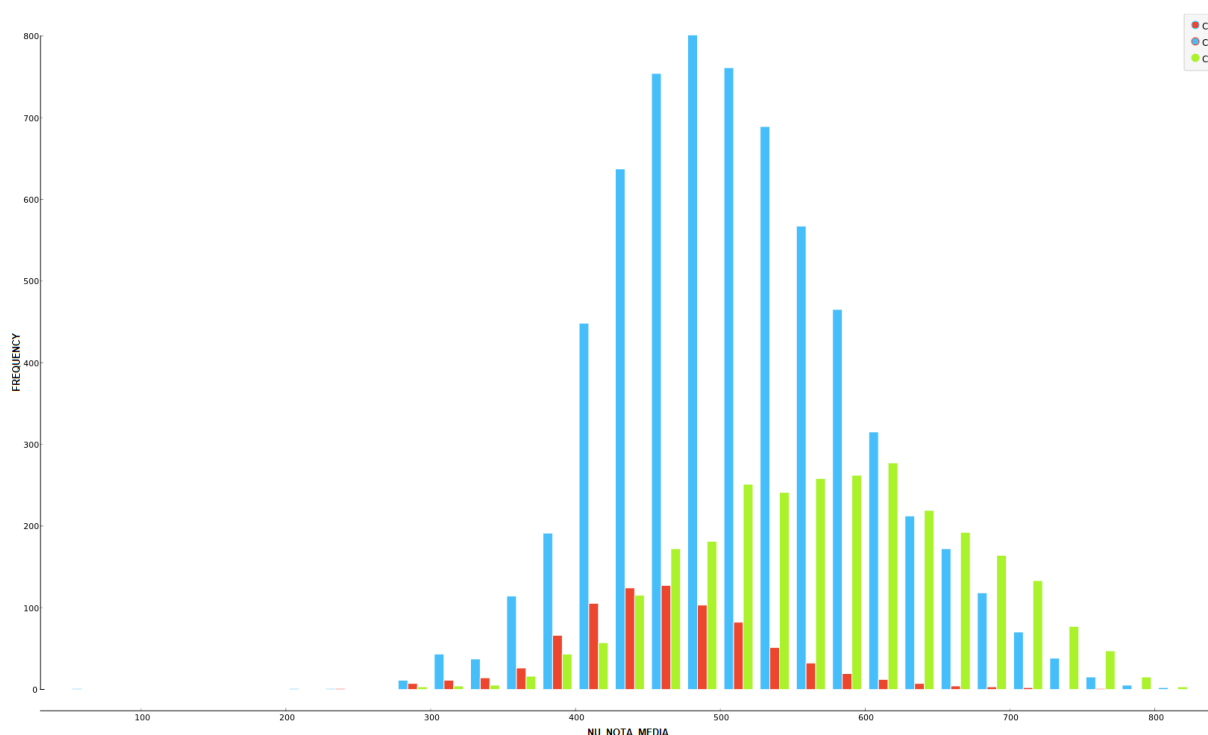
Menos favorecidos	Parcialmente favorecidos	Mais favorecidos
Grupo 1 (7,97%)	Grupo 2 (64%)	Grupo 3 (28,03%)
98% tem renda baixa	75% tem renda baixa	71% tem renda alta ou média
91% não tem nenhum responsável com ensino superior	100% não tem nenhum responsável com ensino superior	100% tem o pai ou a mãe com ensino superior
100% não tem acesso a nenhum tipo de tecnologia.	98% tem acesso a algum tipo de tecnologia	100% tem acesso a algum tipo de tecnologia

Tabela 13 – Agrupamento hierárquico do ENEM de 2020.

A Figura 22 apresenta a distribuição dos participantes de cada grupo de acordo com a nota média dos participantes daquele grupo. Assim como nos anos de pré-pandemia, as melhores notas médias estão no grupo mais favorecido e as piores estão no grupo menos

favorecido, ou seja, no grupo 1. Pode-se notar que quando o valor da nota ultrapassa os 600 pontos, o grupo 3, apesar de ter menos da metade de participantes do grupo 2, tem mais participantes distribuídos naquela faixa de nota do que o grupo 2. É possível notar que a escala muda no ano de 2020, tendo mais participantes com notas entre 100 e 300 pontos, valor que não aparece nos anos de pré-pandemia. Esse resultado repete o comportamento do gráfico de 2019, mostrando que na pandemia, os resultados de cada grupo são iguais os de pré-pandemia.

Figura 22 – Frequência dos participantes de cada grupo em relação à nota média do ENEM no ano de 2020.



Fonte: Elaborado pelo autor (2023).

Os resultados do agrupamento de 2021 estão apresentados na Tabela 14. O grupo 1, ou seja, o menos favorecido, possui a menor porcentagem da amostra em relação aos anos analisados, com cerca de 4,52% do valor total da amostra e com valor menor que o grupo 1 de 2017, 2018, 2019 e 2021. O grupo 2, referente aos favorecidos parcialmente, tem 81,23% da amostra, valor maior que 2020, mas menor que 2019. O grupo mais favorecido teve 14,25% do total da amostra, valor menor que 2020, mas maior que em todos os anos de pré-pandemia analisados. Entretanto, o grupo 2, parcialmente favorecido, tem 37% dos participantes com alta ou média renda, número maior que em todos os anos analisados. Tanto em 2020 quanto em 2021, o grupo mais favorecido teve uma quantidade maior de participantes que nos anos de pré-pandemia. O grupo menos favorecido em 2020 e 2021 foi menor que anos analisados anteriormente, ou seja, durante a pandemia, existiu mais participantes parcialmente favorecidos e mais favorecidos.

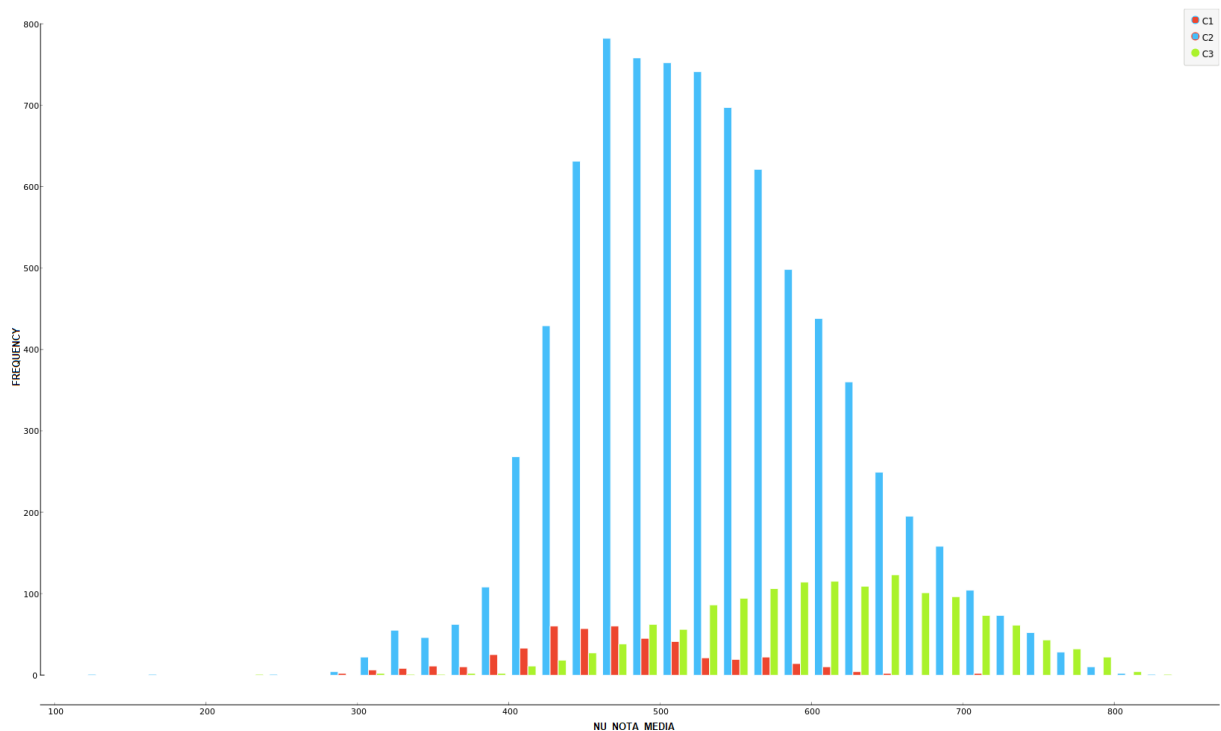
A Figura 23 apresenta a distribuição dos participantes de cada cluster em relação

Menos favorecidos	Parcialmente favorecidos	Mais favorecidos
Grupo 1 (4,52%)	Grupo 2 (81,23%)	Grupo 3 (14,25%)
98% tem renda baixa	63% tem renda baixa	90% tem renda alta ou média
92% não tem nenhum responsável com ensino superior	74% não tem nenhum responsável com ensino superior	100% tem o pai ou a mãe com ensino superior
100% não tem acesso a nenhum tipo de tecnologia.	100% tem acesso a algum tipo de tecnologia	100% tem acesso a algum tipo de tecnologia

Tabela 14 – Agrupamento hierárquico do ENEM de 2021.

à nota média. Assim, como nos anos anteriores, o desempenho dos participantes é maior no grupo mais favorecido e menor no grupo menos favorecido. As variáveis utilizadas para formar os grupos possuem influência na nota dos participantes. Dessa forma, os resultados apresentados pelo gráfico corroboram com os já apresentados nas questões de pesquisa anteriores, ou seja, quanto melhor é a condição social e financeira que o participante se encontra, melhor é o desempenho dele no exame.

Figura 23 – Frequência dos participantes de cada grupo em relação à nota média do ENEM no ano de 2021.



Fonte: Elaborado pelo autor (2023).

## 6 Discussão

Neste capítulo de discussão, são apresentadas as principais conclusões e reflexões que emergiram ao longo da pesquisa, a fim de oferecer uma visão ampla e aprofundada sobre o tema abordado. Além disso, é apresentada a verificação dos critérios de sucesso e, por fim, são discutidas as principais limitações e desafios encontrados ao longo do processo de pesquisa.

### 6.1 Principais Achados

Podemos perceber que, em relação à quantidade de participantes inscritos, ausentes e presentes, a pandemia da COVID-19 impactou de forma negativa a participação no principal exame de avaliação de desempenho no ensino médio. Nos anos anteriores a pandemia, se notava um grande número de inscritos, e que estiveram presentes. Nos anos de 2017, 2018 e 2019, a quantidade de presentes no exame ultrapassava os 3 milhões, e mais de 4 milhões de participantes em 2017. A pandemia no seu primeiro ano fez com que houvesse a maior taxa de abstenções na história do exame, e no segundo ano da pandemia uma baixa quantidade de inscritos comparado aos anos anteriores. Vale ressaltar que, no segundo ano de pandemia, o número de inscritos não alcançou sequer o número de presentes de 2020. A quantidade de inscritos no ano de 2021 foi a menor desde 2005.

Apesar do período pandêmico, os participantes do [ENEM](#) se adaptaram, tentando manter o distanciamento social. Entretanto, uma hipótese acerca do impacto negativo em relação ao número de participantes é que, no ano de 2020, ainda se conhecia pouco sobre a pandemia. Insegurança, alteração das datas das provas, falta de preparação dos candidatos, medo de contaminação da COVID-19 e preocupações financeiras entre os participantes provavelmente contribuíram para o alto número de abstenções em 2020. Uma hipótese para explicar as baixas inscrições do [ENEM](#) de 2021 pode ser devido a muitos transtornos e mudanças na vida das pessoas em relação à educação. O fechamento das escolas e a adoção dos métodos de [EAD](#) podem ter feitos os participantes enfrentarem desafios de aplicação relacionados a acesso à tecnologia ou problemas financeiros causados, fazendo com que os participantes não se sentissem aptos a realizar o exame.

Durante a pandemia de COVID-19, as escolas enfrentaram muitos desafios para continuar a oferecer educação de qualidade aos seus alunos. Algumas escolas foram mais bem equipadas para lidar com esses desafios, enquanto outras enfrentam problemas para se adaptar a esse novo modelo a distância. As escolas de dependência administrativa pública (i.e., estaduais, municipais e federais) enfrentaram maiores dificuldades do que as escolas privadas, devido às diferenças na disponibilidade de recursos e na capacidade de

se adaptar às mudanças, conforme apresentado no Capítulo 2. Entretanto, as análises da nota média e o tipo de dependência administrativa apresentadas no Capítulo 5 mostraram que, durante os anos analisados, não houve alteração significativa no desempenho dos participantes devido ao tipo de dependência administrativa da escola. As Figuras 6 e 7 mostram que as escolas particulares obtiveram melhores resultados nas provas em relação aos outros 3 tipos de dependência administrativa, um resultado já vistos em anos anteriores a pandemia. Dentre as escolas públicas, as escolas federais foram a que obtiveram melhor desempenho ao longo dos anos, seguidas pelas escolas estaduais e municipais. O resultado da questão de pesquisa 2 mostra que a pandemia da COVID-19 não impactou no resultado do ENEM, independente do tipo administrativo da escola. Entretanto, o resultado mostra a necessidade de um melhor investimento em instituições públicas de ensino básico, principalmente de nível estadual e municipal, visto que o desempenho dos participantes dessas escolas é inferior aos demais.

Os resultados obtidos pela questão de pesquisa 3 (QP3), a qual apresenta a nota média geral dos candidatos, conforme o tipo de renda familiar, corrobora com os resultados encontrados na questão de pesquisa 2 (QP2), mostrando que não houve impacto negativo no desempenho dos participantes do ENEM durante a pandemia. Pode-se observar uma disparidade das notas médias entre as diferentes classes de renda familiar. Os resultados mostram que existe um aumento gradativo na média dos grupos de acordo com sua renda familiar, ou seja, quando maior é a renda, maior é a nota média.

Os impactos na nota geral do ENEM ficaram refletidas também nas áreas de conhecimento. Durante o período de pandemia, as notas gerais do ENEM aumentaram, conforme apresentado nas questões de pesquisa 2 e 3. Em 2020, apenas as áreas de matemática e redação não sofreram aumento em relação a 2019. Entretanto, essa diminuição na nota das áreas não impactou na média geral. Em 2021, as médias das áreas voltaram a subir. Foi possível notar que área de Linguagens, códigos e suas naturezas teve menor quantidade de *outliers* negativos em 2021 quando comparado a todos os anos analisados. Os resultados da questão de pesquisa 4, apresentados no capítulo de 5, mostraram impacto positivos no desempenho durante o período da pandemia da COVID-19.

Acerca dos resultados obtidos sobre as maiores correlações, em ambos os anos foi possível notar que a renda familiar possui a maior correlação com a nota média dos participantes. As variáveis socioeconômicas de maior correlação são relacionadas a bens financeiros. Isso nos leva a acreditar que, durante o período pandêmico, os participantes com maiores recursos tecnológicos, como internet, computador e celular, conseguiram aderir melhor ao modelo de ensino à distância. Dessa forma, eles puderam ter facilidade em adquirir o conhecimento necessário para realizar o exame. Essas correlações mais altas em cenários em que os participantes aparentam estar em um nível social-econômico mais alto nos leva a acreditar que a educação ao nível médio difere dependendo do contexto

social em que o aluno está inserido.

Os resultados obtidos para os estados e regiões mostraram que o norte e nordeste tanto antes quanto durante a pandemia tiveram desempenho inferior as demais regiões. Com o aumento do desempenho no ano de 2020 e 2021, quase todos os estados tiveram aumento na nota média geral. Os estados do Maranhão, Amapá, Amazonas e Acre mantiveram as piores colocações dentre os estados em ambos os anos analisados, tanto antes quanto durante a pandemia. Os resultados também mostraram que não houve impacto negativo no desempenho dos estudantes nos anos da pandemia. Como já apresentado anteriormente, o desempenho dos participantes do ENEM aumentou gradativamente em ambos os anos. Alguns estados da região nordeste se sobressaem a média geral da região. Ceará, Rio Grande do Norte e Piauí possuem notas altas, principalmente na área de redação. Durante a pandemia, esses estados tiveram também aumento no desempenho.

Utilizando os 5 anos com uma amostragem de 10.000 participantes, pode-se notar que ao utilizar o modelo de agrupamento hierárquico, os participantes foram divididos em 3 grupos. Após a análise dos grupos, foram criadas as nomenclaturas “mais favorecidos”, “parcialmente favorecidos” e “menos favorecidos”. Quando analisados os anos de 2017, 2018 e 2019, ou seja, a pré-pandemia, foi possível notar que os grupos possuíam uma quantidade de amostras para cada grupo muito parecido, sendo a maior quantidade de participantes alocados no grupo 2, seguido dos participantes do grupo 1 e 3. As características de cada grupo possuíam semelhanças, tornando a análise e comparação de cada um dos grupos de uma forma igual para cada ano, visto que sempre existia um grupo semelhante no ano seguinte. Nos anos de pandemia, o grupo dos menos favorecidos foi menor em ambos os anos, com 7,97% em 2020 e 4,52% em 2021. Em 2019, o grupo dos mais favorecidos teve 8,69% e em 2020 teve 28%, e em 2021 teve 14% aproximadamente. O gráfico de distribuição dos grupos em relação à nota média mostra que os participantes do grupo dos mais favorecidos têm notas mais altas que os demais em ambos os anos analisados. Esse resultado mostra que houve mudanças nas características dos participantes durante o período pandêmico, tendo mais participantes nos grupos mais favorecidos em ambos os anos de pandemia. Isso nos leva a acreditar que essa mudança nas características dos participantes é um dos motivos que pode ter acarretado o leve aumento da nota nos anos de pandemia.

## 6.2 Verificação dos Critérios de Sucesso

Os Critérios de Sucesso desta pesquisa foram, ou estão sendo, alcançados de acordo com o resultado abaixo:

- (CS1) Nesta pesquisa todas as questões de pesquisa foram respondidas e o objetivo de identificar os impactos da pandemia foram alcançados;

- (CS2) Os resultados encontrados mostraram os impactos da pandemia para os estudantes de nível médio. Dentre os principais achados, destaca-se a influência de fatores socioeconômicos, como renda familiar, sobre o desempenho dos estudantes no [ENEM](#) e a abstenções no exame durante a pandemia.
- (CS3) Foram publicados três artigos baseado nesta pesquisa de mestrado, sendo um deles apresentados no Congresso da Escola Regional de Computação do Ceará, Maranhão e Piauí. Um artigo está em produção para publicação. Esta dissertação foi escrita apresentando todos os resultados, portanto, esse critério de sucesso foi alcançado.

### 6.3 Limitações

A alteração das variáveis dos participantes causada no conjunto de dados do ano de 2020 e 2021 limitou os achados desse trabalho. Além disso, esta pesquisa de mestrado utilizou os dados somente até 2021. Como a pandemia da COVID-19 se estendeu por mais um ano e os prejuízos podem aparecer somente a longo prazo, esse estudo acaba não apresentando todas as consequências e impactos causadas pela COVID-19.

Outra limitação deste estudo é o fato de que o [ENEM](#) não dá um panorama exato do ano em questão, já que a prova inclui a participação de alunos já concluintes do ensino médio. Isso pode afetar a interpretação dos resultados, já que a amostra de estudantes que realizam a prova pode não representar totalmente a realidade do ano em questão. Portanto, é importante considerar que os resultados deste estudo podem estar sujeitos a esse viés de amostragem e que é necessário ter cautela ao generalizar os resultados para toda a população de estudantes do ensino médio.

### 6.4 Trabalhos Futuros

A partir desta dissertação de mestrado, é possível realizar outros estudos. Como trabalhos futuros, podemos destacar:

1. Pode ser realizada uma revisão sistemática da literatura sobre a utilização de mineração de dados educacionais durante o período da pandemia. Uma revisão sistemática utilizando apenas bases de dados da educação brasileira durante a pandemia é encorajada;
2. Devido ao fato da pandemia ocorrer até 2023 de acordo com a [OMS](#), é possível realizar uma pesquisa já utilizando os microdados do [ENEM 2023](#) assim que o [INEP](#) fornecer a nova base;

3. Pode-se realizar pesquisas com um escopo menor, visto que o Brasil é um país demograficamente grande. Nesse sentido, investigar o impacto da pandemia em níveis estaduais ou mesmo municipais são outras sugestões para trabalhos futuros. Por exemplo, seria possível ...Maranhão
4. Por fim,



## 7 Considerações Finais

Esta dissertação teve o objetivo de identificar os impactos da pandemia da COVID-19 no [ENEM](#) através da utilização de técnicas de [MDE](#). Para alcançar esse objetivo, foi realizado diversos estudos sobre o funcionamento do exame, suas principais características e sua importância para a educação básica do Brasil através das bases de dados do [INEP](#).

Este trabalho conseguiu alcançar seus objetivos e critérios de sucesso. Por meio de análises descritivas e exploratórias, foi possível identificar impactos positivos e negativos no exame durante o período da pandemia. O trabalho mostrou os impactos da pandemia no [ENEM](#) através de análises exploratórias descritivas, modelo de correlação e modelo de agrupamento, utilizando os cinco anos analisados de forma comparativa.

Os resultados apresentados contribuem para o melhor entendimento do [ENEM](#). O [ENEM](#) é o principal exame de desempenho da educação básica brasileira e porta de entrada para o ensino superior. A análise dos resultados obtidos através da mineração de dados no [ENEM](#) revelaram aspectos importantes relacionados ao desempenho dos estudantes e o impacto da pandemia da COVID-19. As análises dos modelos de correlação e agrupamento evidenciam a importância de medidas que possam diminuir a desigualdade socioeconômica no acesso à educação e aos recursos necessários para um bom desempenho acadêmico. A disparidade entre os participantes vindos de escolas públicas, que tiveram desempenho inferior em relação aos alunos provenientes de escolas particulares, aponta para a necessidade de políticas públicas que possam melhorar a qualidade do ensino público no país. A pandemia também mostrou que é fundamental garantir a realização do [ENEM](#) de forma segura e justa para todos os estudantes para que, em caso de outros eventos inesperados como a pandemia, os estudantes possam realizar o exame.

### 7.1 Contribuições Científicas

As principais contribuições desta pesquisa são:

1. Identificação de trabalhos relacionados que utilizam de modelos e algoritmos computacionais;
2. Mostrou o impacto da pandemia em relação à quantidade de participantes presentes e ausentes;
3. Identificou se o desempenho dos participantes em relação ao tipo administrativo da sua escola foi impactado pela pandemia da COVID-19;

4. A identificação do impacto da pandemia no desempenho dos participantes conforme o grupo de renda familiar pertencente;
5. Apresentou todas as áreas do conhecimento e identificou os impactos nelas durante a pandemia;
6. A identificação de características socioeconômicas de maior correlação com a nota do [ENEM](#);
7. Apresentou os impactos da pandemia no desempenho dos participantes de acordo com seu estado e região;
8. Evidenciou mudanças nas características socioeconômicas dos participantes do [ENEM](#) durante o período pandêmico;
9. Utilização de um modelo de aprendizado máquina em um ambiente educacional;
10. Por fim, a pesquisa utilizou a metodologia [CRISP-DM](#), a última fase consiste na distribuição dos resultados da mineração de dados. Além desta dissertação, os resultados desse estudo levaram a publicações de artigos em revistas e anais, como descrito a seguir.

## 7.2 Publicações

Nessa seção, são apresentadas as publicações decorrentes da pesquisa, destacando-se suas contribuições e importância para a área de Ciência da Computação, aplicada a educação.

1. A primeira publicação a partir desse estudo foi o artigo *A Pandemia da COVID-19 impactou o ENEM? Uma Análise Comparativa de Dados dos Anos de 2019 e 2020*, através da mineração dos dados do [ENEM](#) disponibilizados pelo [INEP](#). O artigo teve o objetivo de comparar os anos de 2019 e 2020 e identificar os impactos da pandemia. Esse artigo foi publicado na Revista Novas Tecnologias na Educação ([NETO et al., 2022b](#));
2. O artigo *Análise Exploratória de Dados para Identificar o Impacto da Pandemia da COVID-19 no ENEM dos Estados do Ceará, Maranhão e Piauí* utilizou os dados de 2019 e 2020, mas analisando detalhadamente e focando nos estados do Maranhão, Ceará e Piauí. Este artigo foi apresentado na 10ª edição da Escola Regional de Computação do Ceará, Maranhão e Piauí (ERCEMAPI) ([NETO et al., 2022a](#));
3. O artigo *Data Analysis To Identify The Impact Of The Covid-19 Pandemic On Enem In 3 States Of Northeast Brazil* estende a análise de dados dos estados do

Maranhão, Ceará e Piauí para os anos de 2017, 2018, 2019, 2020 e 2021. Este artigo foi publicado na Revista de Sistemas e Computação (NETO *et al.*, 2023);

4. O artigo *How did the COVID-19 pandemic affect the performance of high school students in the Brazilian national exam? Findings and Implications through Data Mining*, que analisa o impacto da pandemia, está em produção e para ser submetido na revista Education and Information Technologies.
5. Dois artigos publicados em parceria com outro aluno do Programa de Pós-graduação em Ciência da Computação (PPGCC). Uma revisão sistemática da literatura sobre o uso da mineração de dados usando quatro bases de dados fornecidas pelo INEP (e.g, Sistema de Avaliação da Educação Básica (SAEB), Índice de Desenvolvimento da Educação Básica (IDEB), Indicadores Educacionais e Censo Escolar). Esta revisão foi publicada na Revista Novas Tecnologias na Educação (SOARES *et al.*, 2021); e um artigo sobre os Fatores de Influência da Qualidade Educacional do Maranhão, atualmente em revisão na Revista Brasileira de Informática na Educação.

## Referências

- AHER, S. B.; L.M.R.J., L. Data mining in educational system using weka. *IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT)*, n. 3, p. 20–25, 2011. Citado na página 26.
- ALVES, R. D.; CECHINEL, C.; QUEIROGA, E. Predição do desempenho de matemática e suas tecnologias do ENEM utilizando técnicas de mineração de dados. In: *Workshops do Congresso Brasileiro de Informatica na educação*. Brazilian Computer Society (Sociedade Brasileira de Computação - SBC), 2018. Disponível em: <<https://doi.org/10.5753/cbie.wcbie.2018.469>>. Citado na página 31.
- ALVES, R. D. *et al.* Predição do desempenho da redação do enem utilizando técnicas de mineração de dados. Universidade Federal de Santa Catarina, 2018. Citado na página 31.
- AMARAL, F. *Introdução à Ciência de Dados: Mineração de dados e big data*. [S.l.]: Alta Books Editora, 2016. ISBN 9788576089346. Citado 2 vezes nas páginas 26 e 27.
- ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. Teoria da resposta ao item: conceitos e aplicações. *ABE, Sao Paulo*, 2000. Citado na página 25.
- ANDRADE, T. L. de; RIGO, S. J.; BARBOSA, J. L. V. Active methodology, educational data mining and learning analytics: A systematic mapping study. *Informatics in Education*, Vilnius University Institute of Data Science and Digital Technologies, v. 20, n. 2, p. 171–204, 2021. Citado na página 15.
- AQUINO, E. M. L. *et al.* Medidas de distanciamento social no controle da pandemia de COVID-19: potenciais impactos e desafios no brasil. *Ciência & Saúde Coletiva*, FapUNIFESP (SciELO), v. 25, n. suppl 1, p. 2423–2446, jun. 2020. Disponível em: <<https://doi.org/10.1590/1413-81232020256.1.10502020>>. Citado na página 43.
- ARAÚJO, E. S. C.; SILVA, H. O. d. M. Aplicação de mineração de dados na descoberta dos fatores socioeconômicos associados com o desempenho dos participantes do enem. 2020. Citado na página 31.
- ARRUABARRENA, R. *et al.* On evaluating adaptive systems for education. In: . [S.l.: s.n.], 2002. v. 2347, p. 363–367. ISBN 978-3-540-43737-6. Citado na página 29.
- AZEVEDO, A.; SANTOS, M. Kdd, semma and crisp-dm: A parallel overview. p. 182–185, 01 2008. Citado na página 34.
- BAKER, R. Data mining for education. *International Encyclopedia of Education*, v. 7, p. 112–118, 01 2010. Citado 3 vezes nas páginas 16, 28 e 29.
- BARCELLOS, A. A. *et al.* 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. Citado na página 30.
- BARRETO, A. C. F.; ROCHA, D. S. Covid 19 e educação: resistências, desafios e (im)possibilidades. *Revista Encantar*, v. 2, p. 01–11, maio 2020. Citado na página 22.

- BERKHIN, P. A survey of clustering data mining techniques. In: *Grouping multidimensional data*. [S.l.]: Springer, 2006. p. 25–71. Citado na página 28.
- BERTUCCI-MARTINS, L. M. “conselhos ao povo”: educação contra a influenza de 1918. *Cadernos CEDES*, FapUNIFESP (SciELO), v. 23, n. 59, p. 103–118, abr. 2003. Disponível em: <<https://doi.org/10.1590/s0101-32622003000100008>>. Citado na página 15.
- CAMPOS, R. B. L. *Análise técnica da matriz de referência do ENEM e dos itens de matemática das edições de 2012 a 2014*. Tese (Doutorado) — Programa de Pós-Graduação em Matemática (PROFMAT), 2015. Departamento de Matemática. Disponível em: <<http://www.tede2.ufrpe.br:8080/tede2/handle/tede2/6708>>. Citado 2 vezes nas páginas xi e 25.
- CASTRO, R. P.; AGUILAR, A. G. Factores de éxito de los mooc: algunas consideraciones críticas. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, v. 0, p. 105–118, 03 2014. Disponível em: <<https://doi.org/10.4304/risti.e1.105-118>>. Citado na página 22.
- CHAKRABORTY, I.; MAITY, P. Covid-19 outbreak: Migration, effects on society, global environment and prevention. *Science of The Total Environment*, v. 728, p. 138882, 2020. ISSN 0048-9697. Disponível em: <<https://doi.org/10.1016/j.scitotenv.2020.138882>>. Citado na página 20.
- COLAB. *Google Colaboratory*. 2020. <<https://colab.research.google.com/notebooks/intro.ipynb>>. Citado na página 40.
- DEMŠAR, J. *et al.* Orange: Data mining toolbox in python. *Journal of Machine Learning Research*, v. 14, p. 2349–2353, 2013. Disponível em: <<http://jmlr.org/papers/v14/demsar13a.html>>. Citado na página 40.
- DEMŠAR, J. *et al.* Orange: From experimental machine learning to interactive data mining. In: SPRINGER. *European conference on principles of data mining and knowledge discovery*. [S.l.], 2004. p. 537–539. Citado na página 40.
- DUTT, A. *et al.* Clustering algorithms applied in educational data mining. *International Journal of Information and Electronics Engineering*, IACSIT Press, v. 5, n. 2, p. 112, 2015. Citado na página 27.
- EDUCAÇÃO, T. P. Ensino a distância na educação básica frente à pandemia da covid-19. *Nota Técnica*, 2020. Disponível em: <[https://www.todospelaeducacao.org.br/\\_uploads/\\_posts/425.pdf](https://www.todospelaeducacao.org.br/_uploads/_posts/425.pdf)>. Citado na página 15.
- ELFIRDOUSSI, S. *et al.* Assessing distance learning in higher education during the covid-19 pandemic. *Education Research International*, Hindawi, v. 2020, 2020. Citado na página 21.
- FRANCO, J. *et al.* Usando mineração de dados para identificar fatores mais importantes do enem dos Últimos 22 anos. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 1112–1121. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12867>>. Citado na página 31.
- GHAVAMI, P. *Big Data Analytics Methods*. De Gruyter, 2019. Disponível em: <<https://doi.org/10.1515/9781547401567>>. Citado na página 26.

GOLDSCHMIDT, R.; PASSOS, E. *Data mining*. [S.l.]: Gulf Professional Publishing, 2005. Citado na página 27.

GOMES, R. L. R. e Jorge Luis Cavalcanti Ramos e João Carlos Sedraz Silva e A. S. A literatura brasileira sobre mineração de dados educacionais. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, v. 3, n. 1, p. 621, 2015. ISSN 2316-8889. Disponível em: <<http://dx.doi.org/10.5753/cbie.webie.2014.621>>. Citado na página 33.

INEP. *Relatório Brasil no PISA 2018: Volume I: Competências e desempenho dos estudantes*. [S.l.], 2019. Disponível em: <[https://download.inep.gov.br/publicacoes/institucionais/avaliacoes\\_e\\_examens\\_da\\_educacao\\_basica/relatorio\\_brasil\\_no\\_pisa\\_2018.pdf](https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_examens_da_educacao_basica/relatorio_brasil_no_pisa_2018.pdf)>. Citado na página 16.

INEP. *Censo da Educação Básica 2020*. 2021. <[https://download.inep.gov.br/publicacoes/institucionais/estatisticas\\_e\\_indicadores/notas\\_estatisticas\\_censo\\_escolar\\_2020.pdf](https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/notas_estatisticas_censo_escolar_2020.pdf)>. Citado na página 23.

INEP. *Resposta educacional a pandemia da Covid-19*. 2021. <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/pesquisas-suplementares/pesquisa-resposta-educacional-a-pandemia-de-covid-19>>. Citado na página 22.

INEP. *ENEM Microdados*. 2023. <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>>. Citado na página 37.

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*, v. 18, n. 2, p. 336–345, jan. 2021. Disponível em: <<https://doi.org/10.22456/1679-1916.110252>>. Citado na página 32.

JUNIOR, V. F. d. S. Uma breve história do exame nacional do ensino médio - enem: Avanços e ranços até a era digital / a brief history of the exame nacional do ensino médio - enem (national high school exam): Advances and weaknesses until the digital age. *Brazilian Journal of Development*, v. 7, n. 12, p. 120314–120325, 2021. Disponível em: <<https://doi.org/10.34117/bjdv7n12-683>>. Citado na página 26.

KABAKCHIEVA, D. Student performance prediction by using data mining classification algorithms. *International Journal of Computer Science and Management Research*, v. 1, p. 686–690, 01 2012. Citado na página 16.

KHAN, M. *et al.* Covid-19: A global challenge with old history, epidemiology and progress so far. *Molecules*, v. 26, n. 1, 2021. ISSN 1420-3049. Disponível em: <<https://doi.org/10.3390/molecules26010039>>. Citado 2 vezes nas páginas 15 e 20.

KOMOROWSKI, M. *et al.* Exploratory data analysis. *Secondary Analysis of Electronic Health Records*, Springer International Publishing, p. 185–203, 2016. Citado na página 41.

LIMA, P. d. S. N. *et al.* Análise de dados do enade e enem: uma revisão sistemática da literatura. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, SciELO Brasil, v. 24, p. 89–107, 2019. Citado na página 30.

MARINHO, A. K. B.; BOCHEMBUZIO, C. M. F. Alfabetização e letramento: um olhar crítico sobre o método fônico. *Revista Eletrônica da Educação*, v. 4, n. 2, p. 82–101, 2022. Citado na página 17.

MARTINS, V.; ALMEIDA, J. Educação em tempos de pandemia no brasil: saberesfazeres escolares em exposição nas redes. *Revista Docência e Cibercultura*, Universidade de Estado do Rio de Janeiro, v. 4, n. 2, p. 215–224, ago. 2020. Disponível em: <<https://doi.org/10.12957/redoc.2020.51026>>. Citado na página 22.

MATPLOTLIB. *Matplotlib Documentation*. 2020. <<https://numpy.org/doc/stable/>>. Citado na página 39.

MCKINNEY, W. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. [S.l.]: "O'Reilly Media, Inc.", 2012. Citado na página 39.

MELO, J. E. “seu futuro passa por aqui”: o enem como política avaliativa e os conhecimentos históricos exigidos. *Revista Aedos*, v. 4, n. 11, set. 2012. Citado na página 23.

Ministério da Saúde. *Painel Coronavírus*. 2023. Disponível em <<https://covid.saude.gov.br/>>. Citado na página 15.

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. Citado na página 32.

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. Citado na página 30.

NETO, N. W. *et al.* Análise exploratória de dados para identificar o impacto da pandemia da covid-19 no enem dos estados do ceara, maranhão e piauí. In: *Anais da X Escola Regional de Computação do Ceará, Maranhão e Piauí*. Porto Alegre, RS, Brasil: SBC, 2022. p. 31–40. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/ercemapi/article/view/21957>>. Citado na página 71.

NETO, N. W. *et al.* A pandemia da covid-19 impactou o enem? uma análise comparativa de dados dos anos de 2019 e 2020. *RENOTE*, v. 20, n. 1, p. 223–232, 2022. Disponível em: <<https://seer.ufrgs.br/index.php/renote/article/view/126655>>. Citado na página 71.

NETO, N. W. *et al.* Data analysis to identify the impact of the covid-19 pandemic on enem in 3 states of northeast brazil. *Revista de Sistemas e Computação-RSC*, v. 12, n. 3, 2023. Disponível em: <<https://revistas.unifacs.br/index.php/rsc/article/view/7919>>. Citado na página 72.

NUMPY. *Numpy Documentation*. 2020. <<https://numpy.org/doc/stable/>>. Citado na página 39.

OECD. *Education at a Glance 2021*. [s.n.], 2021. Disponível em: <<https://doi.org/10.1787/b35a14e5-en>>. Citado 2 vezes nas páginas 15 e 16.

PANDAS. *Pandas Documentation*. 2020. <<https://pandas.pydata.org/docs/>>. Citado na página 39.

PIETRA, G. D. *et al.* Por que o brasil vai mal no pisa? uma análise dos determinantes do desempenho no exame. 2018. Citado na página 16.

PIRES, A. Renda familiar e escolaridade dos pais: reflexões a partir dos microdados do enem 2012 do estado de são paulo. *ETD Educação Temática Digital*, v. 17, n. 3, p. 523–541, 2015. Disponível em: <<https://periodicos.sbu.unicamp.br/ojs/index.php/etd/article/view/8638262>>. Citado na página 17.

PRABHA, S. L.; SHANAVAS, A. M. Educational data mining applications. *Operations Research and Applications: An International Journal (ORAJ)*, v. 1, n. 1, p. 23–29, 2014. Citado na página 27.

RAMOS, J. *et al.* Crisp-edm: uma proposta de adaptação do modelo crisp-dm para mineração de dados educacionais. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 1092–1101. ISSN 0000-0000. Disponível em: <<https://doi.org/10.5753/cbie.sbie.2020.1092>>. Citado na página 34.

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. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0738059321001383>>. Citado na página 32.

RIBEIRO, M. X.; VIEIRA, M. T. P.; TRAINA, A. J. M. Mineração de regras de associação usando agrupamentos. In: *I workshop sobre algoritmos de mineração de dados*. [S.l.: s.n.], 2005. p. 9–16. Citado na página 29.

ROBBINS, J. *UNESCO: 290 Million Students Stay Home due to Coronavirus*. 2020. Disponível em: <<https://learningenglish.voanews.com/a/unesco-290-million-students-stay-home-due-to-coronavirus/5317148.html>>. Citado 2 vezes nas páginas 15 e 20.

SAIZ-MANZANARES, M. C. *et al.* Monitoring of student learning in learning management systems: An application of educational data mining techniques. *Applied Sciences*, v. 11, n. 6, 2021. ISSN 2076-3417. Disponível em: <<https://doi.org/10.3390/app11062677>>. Citado na página 32.

SANTANA, C. L. S. e.; SALES, K. M. B. Aula em casa: educação, tecnologias digitais e pandemia covid-19. *Interfaces Científicas - Educação*, v. 10, n. 1, p. 75–92, set. 2020. Disponível em: <<https://periodicos.set.edu.br/educacao/article/view/9181>>. Citado na página 22.

SANTOS, F. de P.; SILVEIRA, I. F.; LECHUGO, C. P. Análise da percepção dos alunos sobre as práticas docentes por meio da mineração de dados educacionais. *Revista Espacios*, v. 38, n. 05, 2017. Citado 2 vezes nas páginas 27 e 28.

SANTOS, M. F. d. S. *et al.* Uma análise do discurso sobre educação publicado no twitter em tempo de covid-19 no brasil. Universidade Federal da Paraíba, 2021. Citado na página 32.

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. Citado 2 vezes nas páginas 34 e 35.



SILVA, L.; MORINO, A.; SATO, T. Prática de mineração de dados no exame nacional do ensino médio. In: . [S.l.: s.n.], 2014. p. 651. Citado na página 30.

SILVA, M. V. e C. A educação no contexto da pandemia de covid-19: uma revisão sistemática de literatura. *Revista Brasileira de Informática na Educação*, v. 28, n. 0, p. 1013–1031, 2020. ISSN 2317-6121. Disponível em: <<http://dx.doi.org/10.5753/rbie.2020.28.0.1013>>. Citado na página 22.

SILVA, V. *et al.* Identificação de desigualdades sociais a partir do desempenho dos alunos do ensino médio no enem 2019 utilizando mineração de dados. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 72–81. ISSN 0000-0000. Disponível em: <<https://doi.org/10.5753/cbie.sbie.2020.72>>. Citado na página 31.

SILVEIRA, F. L. d.; BARBOSA, M. C. B.; SILVA, R. d. Exame nacional do ensino médio (enem): Uma análise crítica. *Revista Brasileira de Ensino de Física*, Sociedade Brasileira de Física, v. 37, n. Rev. Bras. Ensino Fís., 2015 37(1), 2015. ISSN 1806-1117. Disponível em: <<https://doi.org/10.1590/S1806-11173710001>>. Citado na página 23.

SOARES, R. de C. *et al.* Mineracão de dados da educação básica brasileira usando as bases do inep: Uma revisão sistemática da literatura. *RENOTE*, v. 19, n. 1, p. 361–370, jul. 2021. Disponível em: <<https://doi.org/10.22456/1679-1916.118526>>. Citado 2 vezes nas páginas 30 e 72.

SOUSA, J. R. d. *et al.* Python e predição de dados usando redes neurais multicamadas/python and data prediction using multi-layered neural networks. *Brazilian Journal of Development*, v. 6, n. 7, p. 54181–54185, Aug. 2020. Disponível em: <<https://doi.org/10.34117/bjdv6n7-897>>. Citado na página 39.

SRINATH, K. Python—the fastest growing programming language. *International Research Journal of Engineering and Technology*, v. 4, n. 12, p. 354–357, 2017. Citado na página 39.

SUNHARE, P.; CHOWDHARY, R. R.; CHATTOPADHYAY, M. K. Internet of things and data mining: An application oriented survey. *Journal of King Saud University - Computer and Information Sciences*, v. 34, n. 6, Part B, p. 3569–3590, 2022. ISSN 1319-1578. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S131915782030416X>>. Citado na página 26.

SURYASA, I. W.; RODRÍGUEZ-GÁMEZ, M.; KOLDORIS, T. The covid-19 pandemic. *International journal of health sciences*, v. 5, n. 2, p. 6–9, 2021. Disponível em: <<https://sciencescholar.us/journal/index.php/ijhs/article/view/2937>>. Citado na página 15.

TRAVITZKI, R. *ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar*. Tese (Doutorado) — Universidade de São Paulo, Agência USP de Gestao da Informacao Academica (AGUIA), 2013. Disponível em: <<https://doi.org/10.11606/t.48.2013.tde-28062013-162014>>. Citado 3 vezes nas páginas xi, 23 e 24.

TSAI, C.-W. *et al.* Big data analytics: a survey. *Journal of Big Data*, Springer Science and Business Media LLC, v. 2, n. 1, out. 2015. Disponível em: <<https://doi.org/10.1186/s40537-015-0030-3>>. Citado na página 26.

- VASCONCELOS, J. C. *et al.* Infraestrutura escolar e investimentos públicos em educação no brasil: a importância para o desempenho educacional. *Ensaio: Avaliação e Políticas Públicas em Educação*, Fundação CESGRANRIO, v. 29, p. 874–898, 2021. ISSN 0104-4036. Disponível em: <<https://doi.org/10.1590/S0104-40362020002802245>>. Citado na página 17.
- VIEIRA, M. D. F.; SILVA, C. M. S. da. A educação no contexto da pandemia de COVID-19: uma revisão sistemática de literatura. *Revista Brasileira de Informática na Educação*, Sociedade Brasileira de Computacao - SB, v. 28, p. 1013–1031, fev. 2020. Disponível em: <<https://doi.org/10.5753/rbie.2020.28.0.1013>>. Citado na página 20.
- VIGGIANO, E.; MATTOS, C. O desempenho de estudantes no enem 2010 em diferentes regiões brasileiras. *Revista Brasileira de Estudos Pedagógicos*, v. 94, 06 2019. Citado na página 23.
- VIRTANEN, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, Springer Science and Business Media LLC, v. 17, n. 3, p. 261–272, 2020. Disponível em: <<https://doi.org/10.1038/s41592-019-0686-2>>. Citado na página 39.
- WASKOM, M. L. seaborn: statistical data visualization. *Journal of Open Source Software*, The Open Journal, v. 6, n. 60, p. 3021, 2021. Disponível em: <<https://doi.org/10.21105/joss.03021>>. Citado na página 39.
- ZHAI, P. *et al.* The epidemiology, diagnosis and treatment of covid-19. *International Journal of Antimicrobial Agents*, v. 55, n. 5, p. 105955, 2020. ISSN 0924-8579. Disponível em: <<https://doi.org/10.1016/j.ijantimicag.2020.105955>>. Citado na página 20.