



UNIVERSIDADE DE BRASÍLIA  
FACULDADE DE CIÊNCIA DA INFORMAÇÃO  
CURSO DE GRADUAÇÃO EM BIBLIOTECONOMIA

GUILHERME COUTO DOS SANTOS

**Dados do Exame Nacional do Ensino Médio (ENEM):  
análise dos perfis socioeconômicos e demográficos**

Brasília  
2025

GUILHERME COUTO DOS SANTOS

**Dados do Exame Nacional do Ensino Médio (ENEM):  
análise dos perfis socioeconômicos e demográficos**

Monografia submetida ao Curso de Graduação em Biblioteconomia da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Biblioteconomia.

Orientador: Prof. Dr. Márcio de Carvalho Victorino

Brasília  
2025

**FOLHA DE APROVAÇÃO**

**Título:** Estudo dos perfis socioeconômicos e demográficos dos participantes do Exame Nacional do Ensino Médio (ENEM) utilizando a análise dos microdados fornecidos pelo INEP.

**Autor(a):** Guilherme Couto dos Santos

Monografia apresentada em **18 de fevereiro de 2025** à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Orientador(a) (FCI/UnB): Dr. Márcio de Carvalho Victorino

Membro Interno (FCI/UnB): Dr. Felipe Augusto Arakaki

Membro Externo (IPHAN): Ma. Jéssica Bilac Gaspareto



Documento assinado eletronicamente por **Jessica Bilac Gaspareto, Usuário Externo**, em 18/03/2025, às 22:51, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Felipe Augusto Arakaki, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 19/03/2025, às 22:12, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.

---



Documento assinado eletronicamente por **Marcio de Carvalho Victorino, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 21/03/2025, às 11:40, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.

---



A autenticidade deste documento pode ser conferida no site  
[http://sei.unb.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0),  
informando o código verificador **12541016** e o código CRC **02FD275C**.

---

**Referência:** Processo nº 23106.029556/2025-61

SEI nº 12541016

### CIP - Catalogação na Publicação

CC871d Couto, Guilherme.  
Dados do Exame Nacional do Ensino Médio (ENEM): análise dos perfis socioeconômicos e demográficos / Guilherme Couto;  
Orientador: Márcio de Carvalho Victorino. -- Brasília, 2025.  
70 f.

Trabalho de Conclusão de Curso (Graduação - Biblioteconomia) -- aqui Universidade de Brasília, 2025.

1. Microdados. 2. Exame Nacional do Ensino Médio. 3. Educação. I. de Carvalho Victorino, Márcio, orient. II. Título.

Dedico este trabalho à minha família. Se cheguei até aqui, foi por vocês.

## **Agradecimentos**

A realização deste trabalho só foi possível graças ao apoio e colaboração de muitas pessoas, às quais expresso minha gratidão.

Primeiramente, agradeço ao meu orientador, Prof. Dr. Márcio de Carvalho Victorino por sua dedicação, paciência e valiosas orientações ao longo de todo o processo de desenvolvimento deste trabalho. Sua experiência e incentivo foram essenciais para a conclusão desta pesquisa. Também agradeço a ele por despertar meu interesse na área de tecnologia, me mostrando o potencial do bibliotecário.

À minha família, por ter me apoiado e investido na construção de minha educação.

Aos meus amigos e colegas de curso, pelo companheirismo e pelas discussões construtivas que enriqueceram meu aprendizado.

À minha companheira por sempre estar ao meu lado.

Quero agradecer aos professores da Faculdade de Ciência da Informação por contribuírem com minha formação acadêmica e profissional ao longo desses anos.

Por fim, agradeço a todos que, direta ou indiretamente, colaboraram para que este trabalho fosse concluído, incluindo aqueles que participaram de discussões, ofereceram conselhos ou forneceram inspiração.

A todos vocês, minha mais sincera gratidão.

*“Só por que alguma coisa não faz o que você planejou que ela fizesse não quer dizer que ela seja inútil”.*

Thomas Edison



## Resumo

O estudo se insere no contexto da análise de dados educacionais e socioeconômicos, buscando compreender as disparidades no acesso à educação superior no Brasil, promover a inclusão e fornecer informações para o desenvolvimento de políticas públicas voltadas à equidade e à redução das desigualdades educacionais. Com isso, o objetivo principal é organizar os dados disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) sobre o Exame Nacional do Ensino Médio (ENEM) 2023 para proporcionar a análise das características socioeconômicas e demográficas predominantes dos participantes desse exame, com o intuito de subsidiar decisões do Ministério da Educação (MEC).

Utilizou-se uma abordagem quantitativa com a análise dos microdados do ENEM 2023. Esses dados foram organizados em tabelas, contendo informações sobre provas, gabaritos, notas e questionários dos participantes. A análise inicial consistiu na seleção dos dados relevantes, seguida do tratamento quantitativo utilizando a linguagem de programação Python, com as bibliotecas pandas e matplotlib. A visualização dos dados foi realizada por meio de gráficos gerados no Jupyter Notebooks.

A análise dos microdados revelou que a maioria dos participantes tem 18 anos, reforçando o ENEM como um importante acesso ao ensino superior. Observou-se também uma predominância de participantes mulheres, especialmente no Nordeste, e uma alta concentração de estudantes provenientes de escolas públicas. A maioria dos inscritos se autodeclarou como parda, destacando a diversidade racial no exame. Dados sobre a escolaridade dos pais indicam que muitos candidatos vêm de famílias com ensino médio completo, mas sem acesso ao ensino superior, o que reflete as desigualdades no progresso educacional. Além disso, a maioria das famílias tem uma renda mensal de até R\$ 1.320,00 e a falta de recursos tecnológicos impacta negativamente a preparação dos candidatos.

Este estudo contribui para a compreensão das disparidades socioeconômicas no acesso ao ensino superior e pode orientar o MEC na formulação de políticas públicas mais inclusivas, com foco na equidade e no combate às desigualdades históricas. O mapeamento das variáveis socioeconômicas dos participantes do ENEM pode ser utilizado como base para a criação de medidas que garantam condições mais justas para todos os candidatos.

**Palavras-chave:** microdados, ENEM, socioeconômico, educação, políticas públicas.

## Abstract

The study is set within the context of the analysis of educational and socioeconomic data, aiming to understand the disparities in access to higher education in Brazil, promote inclusion, and provide information for the development of public policies focused on equity and reducing educational inequalities. The main objective is to organize the data provided by the Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) regarding the 2023 Exame Nacional do Ensino Médio (ENEM) in order to analyze the predominant socioeconomic and demographic characteristics of the exam participants, with the purpose of supporting decisions by the Ministério da Educação (MEC).

A quantitative approach was used to analyze the 2023 ENEM microdata. These data were organized in tables containing information about the exams, answer sheets, scores, and participant questionnaires. The initial analysis involved selecting relevant data, followed by quantitative processing using the Python programming language, with the pandas and matplotlib libraries. Data visualization was performed through graphs generated in Jupyter Notebooks.

The analysis of the microdata revealed that the majority of participants are 18 years old, reinforcing ENEM as an important access point to higher education. It was also observed that there is a predominance of female participants, especially in the Northeast, and a high concentration of students from public schools. Most of the participants self-identified as brown, highlighting the racial diversity in the exam. Data regarding the education level of parents indicate that many candidates come from families with high school education but without access to higher education, reflecting inequalities in educational progress. Furthermore, most families have a monthly income of up to R\$ 1,320.00, and the lack of technological resources negatively impacts candidates' preparation.

This study contributes to the understanding of socioeconomic disparities in access to higher education and can guide the MEC in formulating more inclusive public policies, focusing on equity and addressing historical inequalities. Mapping the socioeconomic variables of ENEM participants can serve as the basis for creating measures to ensure fairer conditions for all candidates.

**Keywords:** microdata, ENEM, socioeconomic, education, public policies.

## **Lista de ilustrações**

Figura 1 – Evolução do número de inscritos . . . . .	34
Figura 2 – Arquitetura da Informação da Análise . . . . .	51
Figura 3 – Gráfico da distribuição de faixa etária . . . . .	53
Figura 4 – Gráfico da distribuição por regiões do Brasil . . . . .	54
Figura 5 – Gráfico da distribuição por regiões do Brasil e sexo . . . . .	55
Figura 6 – Gráfico da distribuição por tipo de escola do Ensino Médio . . . . .	56
Figura 7 – Gráfico da distribuição por cor ou raça . . . . .	57
Figura 8 – Gráfico da distribuição do número de computadores por residência . . . . .	58
Figura 9 – Gráfico da distribuição do acesso à Internet . . . . .	59

## Lista de tabelas

Tabela 1 – Formato de dados abertos . . . . .	23
Tabela 2 – Tipos de uso . . . . .	24
Tabela 3 – Princípios Básicos dos Dados Abertos Governamentais . . . . .	25
Tabela 4 – Licenças abertas . . . . .	26
Tabela 5 – Principais Elementos do ODIN . . . . .	29
Tabela 6 – Percurso histórico do ENEM . . . . .	33
Tabela 7 – TP_FAIXA_ETARIA . . . . .	43
Tabela 8 – TP_SEXO . . . . .	43
Tabela 9 – TP_COR_RACA . . . . .	44
Tabela 10 – TP_ESCOLA . . . . .	44
Tabela 11 – SG_UF_PROVA . . . . .	45
Tabela 12 – Q001 . . . . .	46
Tabela 13 – Q002 . . . . .	47
Tabela 14 – Q006 . . . . .	48
Tabela 15 – Q024 . . . . .	48
Tabela 16 – Q025 . . . . .	49
Tabela 17 – Distribuição da Escolaridade do Pai . . . . .	60
Tabela 18 – Distribuição da Escolaridade da Mãe . . . . .	61
Tabela 19 – Distribuição da Renda Familiar . . . . .	62

### **Lista de abreviaturas e siglas**

BNCC	Base Nacional Comum Curricular
BNCC	Base Nacional Comum Curricular
CGU	Controladoria-Geral da União
CI	Ciência da Informação
DGAs	Dados Governamentais Abertos
ENCCEJA	Exame Nacional de Certificação da Educação de Jovens e Adultos
ENEM	Exame Nacional do Ensino Médio
FIES	Fundo de Financiamento Estudantil
INDA	Infraestrutura Nacional de Dados Abertos
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
LAI	Lei de Acesso à Informação
LGPD	Lei Geral de Proteção de Dados
LSI	Indexação Semântica Latente
MEC	Ministério da Educação
OC	Organização do Conhecimento
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
OD	Open Definition
OD4D	Open Data for Development
ODB	Open Data Barometer
ODIN	Open Data Inventory
OGP	Open Government Partnership
OI	Organização da Informação
PDA	Política de Dados Abertos
PDDL	Open Data Commons Public Domain Dedication and License
PROUNI	Programa Universidade para Todos

RI	Recuperação da Informação
SISU	Sistema de Seleção Unificada
SRI	Sistemas de Informação
TFD	Teoria Fundamentada nos Dados
TREC	Text Retrieval Conference
e-SIC	Sistema Eletrônico do Serviço de Informações ao Cidadão
ePING	Padrões de Interoperabilidade de Governo Eletrônico

## Sumário

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>12</b>
<b>1.1</b>	<b>Problema de Pesquisa . . . . .</b>	<b>13</b>
<b>1.2</b>	<b>Justificativa . . . . .</b>	<b>13</b>
<b>1.3</b>	<b>Objetivo Geral . . . . .</b>	<b>14</b>
<b>1.4</b>	<b>Objetivos específicos . . . . .</b>	<b>14</b>
<b>2</b>	<b>REVISÃO DE LITERATURA . . . . .</b>	<b>15</b>
<b>2.1</b>	<b>CIÊNCIA DA INFORMAÇÃO . . . . .</b>	<b>15</b>
<b>2.2</b>	<b>GESTÃO DA INFORMAÇÃO . . . . .</b>	<b>19</b>
<b>2.3</b>	<b>DADOS ABERTOS . . . . .</b>	<b>20</b>
2.3.1	Open Government Partnership . . . . .	26
2.3.2	Open Data Barometer . . . . .	27
2.3.3	Open Data Inventory . . . . .	28
<b>2.4</b>	<b>DADOS ABERTOS DO INEP . . . . .</b>	<b>30</b>
<b>2.5</b>	<b>ANÁLISE DA INFORMAÇÃO . . . . .</b>	<b>30</b>
<b>2.6</b>	<b>O EXAME NACIONAL DO ENSINO MÉDIO . . . . .</b>	<b>32</b>
<b>3</b>	<b>METODOLOGIA . . . . .</b>	<b>39</b>
<b>4</b>	<b>ANÁLISE DE DADOS . . . . .</b>	<b>40</b>
<b>4.1</b>	<b>Conjunto de Dados . . . . .</b>	<b>40</b>
<b>4.2</b>	<b>Arquitetura da Informação da Análise . . . . .</b>	<b>49</b>
<b>4.3</b>	<b>Elementos Gráficos . . . . .</b>	<b>52</b>
4.3.1	Distribuição de Faixa Etária . . . . .	53
4.3.2	Distribuição por Regiões do Brasil . . . . .	54
4.3.3	Distribuição por Regiões do Brasil e Sexo . . . . .	55
4.3.4	Distribuição por Tipo de Escola do Ensino Médio . . . . .	56
4.3.5	Distribuição por Cor ou Raça . . . . .	57
4.3.6	Distribuição do Número de Computadores Por Residência . . . . .	58
4.3.7	Distribuição do Acesso à Internet . . . . .	59
4.3.8	Distribuição da Escolaridade do Pai . . . . .	60
4.3.9	Distribuição da Escolaridade da Mãe . . . . .	61
4.3.10	Distribuição da Renda Familiar . . . . .	62
<b>5</b>	<b>Conclusão . . . . .</b>	<b>64</b>
<b>6</b>	<b>Referências . . . . .</b>	<b>67</b>

## 1 INTRODUÇÃO

O Ensino Médio brasileiro consiste à etapa do sistema de ensino equivalente à última fase da educação básica, tendo como finalidade, aprofundar os conhecimentos adquiridos no ensino fundamental, a preparação básica do educado para o trabalho e a cidadania, aprimoramento do educado como pessoa humana e a compreensão dos fundamentos científico-tecnológicos dos processos produtivos. A fim de avaliar o desempenho do estudante ao fim de sua escolaridade básica, o Exame Nacional do Ensino Médio (ENEM) foi criado em 1998. Podem participar do exame alunos que concluíram ou que estão concluindo o ensino médio em anos anteriores. Os dados do ENEM são disponibilizados ao público periodicamente, integrando, assim, a iniciativa de dados abertos do governo federal.

A Open Knowledge Foundation, organização sem fins lucrativos que promove o conhecimento livre, define que dados abertos são dados que podem ser usados, compartilhados e desenvolvidos livremente por qualquer pessoa, em qualquer lugar, para qualquer finalidade (James, 2013). Nesse sentido, o Governo Federal do Brasil, na busca por uma administração que valoriza a transparência, tem em sua Constituição Federal (1988) inciso XXXIII do art. 5.º o seguinte:

XXXIII - todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado; (Brasil, 2005).

A lei que garante que se cumpra o inciso supracitado é a Lei N° 12.527, de 18 de novembro de 2011, A Lei de Acesso à Informação (LAI). Essa lei é um marco importante no tocante da disponibilização de informação pelos órgãos e entidades públicas. A LAI prevê que podem solicitar informações pessoas de qualquer idade e nacionalidade, além de empresas e organizações. Dessa forma, a LAI passa a ser um incentivo para a produção, tratamento e organização de dados abertos governamentais, incluindo dados educacionais.

Um exemplo prático disso, é disponibilizado pelo Ministério da Educação (MEC), que divulga as principais bases de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Em 1937 foi criado o INEP, órgão federal originalmente nomeado “Instituto Nacional de Pedagogia” e em 1997, transformado em autarquia federal vinculada ao MEC. “[. . .] é o órgão federal responsável pelas evidências educacionais e atua em três esferas: avaliações e exames educacionais; pesquisas estatísticas e indicadores educacionais; e gestão do conhecimento e estudos.” (INEP, 2024).

O INEP disponibiliza os microdados do ENEM, que consistem no menor nível de desagregação de dados recolhidos por meio do exame. “Eles atendem a demanda por informações específicas ao disponibilizar as provas, os gabaritos, as informações sobre



os itens, as notas e o questionário respondido pelos inscritos no ENEM.” (Ministério da Educação, 2024).

A partir dos dados oferecidos, que contêm informações tanto do desempenho dos participantes quanto de suas características socioeconômicas, o presente trabalho aplica princípios da Ciência da Informação para organizar, explorar e transformar esses dados em conhecimento, proporcionando subsídios às decisões na área de educação.

## **1.1 Problema de Pesquisa**

Como organizar os dados disponibilizados pelo INEP sobre o ENEM para proporcionar a análise das características socioeconômicas e demográficas predominantes dos participantes desse exame a fim de subsidiar decisões do MEC?

## **1.2 Justificativa**

A crescente produção de dados causada pela explosão informacional que aconteceu após o fim da Segunda Guerra Mundial e que, ao passar dos anos desde então, se intensifica com o surgimento e aprimoramento da tecnologia, se faz necessário que além do papel já desempenhado pela Ciência da Informação (CI) na coleta, tratamento, disseminação, armazenamento e recuperação da informação, se utilizem ferramentas e métodos capazes de analisar essas informações proporcionando uma compreensão mais ampla sobre o significado dos dados.

A análise de dados organizados é um processo crítico que fornece indicadores valiosos e conclusões alinhadas com a realidade. Estes dados são derivados de uma variedade de fontes e podem variar em complexidade, mas o objetivo final é sempre o mesmo: fornecer percepções valiosas que possam ser utilizados na tomada de decisões mais assertivas. Através deste processo, é possível identificar tendências, padrões e correlações nos dados que podem não ser imediatamente óbvios.

Entender o perfil dos participantes do ENEM justifica-se por sua relevância para políticas públicas em educação e na demonstração de como a informação pode ser utilizada em prol do desenvolvimento do contexto educacional do país. Assim, é possível notar que a análise desses fatores pode impactar direta ou indiretamente a formulação de estratégias de ensino e políticas de inclusão educacional, por um estudo sistemático. Para tanto, é essencial discorrer sobre a organização da informação, tratar e analisar os microdados do ENEM 2023 por meio de gráficos gerados.

### **1.3 Objetivo Geral**

Organizar os dados disponibilizados pelo INEP sobre o ENEM para proporcionar a análise das características socioeconômicas e demográficas predominantes dos participantes desse exame a fim de subsidiar decisões do MEC.

### **1.4 Objetivos específicos**

- 1) Tratar os microdados coletados do site do INEP referentes ao ENEM 2023 a fim de manter apenas os dados relevantes ao objetivo geral do estudo;
- 2) Gerar gráficos para facilitar a visualização dos dados a fim de subsidiar decisões;
- 3) Analisar os microdados por meio dos gráficos gerados.

## 2 REVISÃO DE LITERATURA

### 2.1 CIÊNCIA DA INFORMAÇÃO

O conceito do termo “informação” é complexo e multifacetado, com definições e limites de seus usos que podem ser difíceis de determinar com clareza. Um dos motivos que justifica a dificuldade em definir este termo, segundo Francelin e Pellegatti (2004, p.124), é que “Numa disposição formal, o fenômeno da informação é estudado em disciplinas diversas, confirmando assim as ramificações complexas e muitas manifestações a ele associadas”.

Buckland (1991) classifica os três principais usos da informação, com o objetivo de trazer luz a essa questão. A informação como processo é a ação de informar, é a aplicação da informação ligada à comunicação e a transmissão da informação. Quando essa transmissão acontece, o indivíduo (interlocutor) não só recebe novos dados, mas modifica seu conhecimento pré-existente. Essa ideia é reforçada por Brookes (1980) que afirma que “A informação é um elemento que promove transformações nas estruturas do indivíduo, sendo essas estruturas de caráter subjetivo ou objetivo”.

O próximo uso descrito é o da informação como conhecimento, é o resultado da ação, o conhecimento transmitido pela ação de comunicar. Os dois primeiros usos são complementares e se interligam, já que a necessidade da informação surge da percepção do próprio indivíduo sobre seu estado de conhecimento:

Nosso estado (ou nossos estados) de conhecimento sobre determinado assunto, em determinado momento, é representado por uma estrutura de conceitos ligados por suas relações: nossa ‘imagem’ do mundo. Quando constatamos uma deficiência ou uma anomalia desse(s) estados(s) de conhecimento encontramos-nos em um estado anômalo de conhecimento. Tentamos obter uma informação ou informações que corrigirão essa anomalia. Disso resultará um novo estado de conhecimento. (Le Coadic, 2002, p. 8).

O conhecimento, internalizado como informação nas estruturas cognitivas do indivíduo, possui a característica da intangibilidade. No entanto, para ser comunicado, o conhecimento precisa ser representado de alguma forma física. Esse conceito é incorporado pela terceira categoria de informação: a informação como coisa. Le Coadic (2002, p.5) afirma categoricamente que “a informação é um conhecimento inscrito (registrado) em forma escrita (impressa ou numérica), oral ou audiovisual”. A construção do conhecimento depende do acesso e processamento da informação, a qual precisa ser organizada e disseminada (Maimone; Silveira, 2007, p. 64).

Por mais que, mesmo dentro da área da CI, não exista uma definição universal do que é a informação, somente a informação como coisa pode passar pelo processo de organização, armazenamento, recuperação e utilização da mesma. A informação como coisa

é um elemento crucial nos estudos de Sistemas de Informação e Ciência da Informação, pois esses campos lidam diretamente com a informação física, independentemente do seu formato ou meio.

A ciência é uma construção social moldada por fatores históricos e socioeconômicos (Le Coadic 2002, p. 19). Concordam Miranda (2002) e outros autores que a origem da CI como disciplina, acontece em decorrência da “explosão informacional” após a Segunda Guerra Mundial. Ortega (2004) complementa afirmando que:

As consequências da Segunda Guerra Mundial, tanto de evolução quanto de reconstrução, e o grande desenvolvimento tecnológico ocorrido a partir dos anos 50 com o surgimento dos computadores figuram como elementos importantes no quadro internacional, influenciando sobre o modo como a ciência e as técnicas relacionadas à organização da informação caminharam em cada cultura. (Ortega, 2004, p.3)

Para Saravevic (1995, p.1), a origem da CI está ligada a um artigo escrito em 1945 por Vannevar Bush, um dos cientistas mais influentes de seu tempo, que chegou a comandar 6000 cientistas. Ele identificou o problema causado pelo aumento exponencial da produção da informação e da ineficiência dos sistemas de produção, organização, acesso, recuperação e disseminação dessa informação:

A dificuldade parece residir não tanto em publicarmos excessivamente em vista da amplitude e variedade dos interesses atuais, mas sim no fato de a publicação ter se expandido muito além da nossa capacidade atual de utilizar efetivamente o registro. (Vannevar, 1945, p. 3, tradução nossa)

Bush (1945) afirmava que os mecanismos de organização da informação adotados pelas bibliotecas à época tinham a problemática de organizar a informação de forma artificial, linear, diferentemente da forma como a mente humana processa e recupera a informação, dificultando esse processo:

Quando dados de qualquer tipo são armazenados, eles são arquivados em ordem alfabética ou numérica, e a informação é encontrada (quando é) seguindo uma busca de subclasse em subclasse. (Vannevar, 1945, p. 14, tradução nossa)

A solução proposta por Vannevar foi uma máquina chamada “memex”, que tem o conceito de memória externa de um indivíduo, capaz de duplicar “artificialmente sua memória”. A visão de Bush é uma previsão da criação da CI, e o início da inteligência artificial e redes neurais.

A CI se consolidou como uma ciência social e interdisciplinar. Social por reconhecer o impacto do problema da informação para o ser social em sua busca por conhecimento (Le Coadic, 2002, p. 21). Interdisciplinar por compreender a necessidade de transcender os

limites de disciplinas isoladas para abordar a complexa natureza da informação (Popper, 1992).

O surgimento do termo “ciência da informação” na década de 1960 (Helprin, 1989, apud Pinheiro; Loureiro, 1995), marcou um momento crucial na história da organização e do acesso ao conhecimento. Essa nova área do saber nasce da questão básica de acesso à informação, ou facilitação desse acesso (Dias, 2001).

A partir da década de 1950, um movimento colaborativo entre cientistas, engenheiros e governos foi iniciado. Dos esforços empenhados para encontrar soluções para a “explosão informacional”, a Recuperação da Informação (RI) despontou como um dos pilares fundamentais da CI. Em 1950, quando Calvin Mooers apresentou um artigo em uma conferência, o termo “recuperação da informação” foi utilizado pela primeira vez:

O problema em discussão aqui é a busca e recuperação de informações por máquinas a partir de um armazenamento conforme uma especificação por assunto [...] não deveria ser necessário enfatizar a importância da recuperação da informação perante um grupo científico como este, pois todos nós já experimentamos a frustração com o funcionamento de nossas bibliotecas – todas as bibliotecas, sem exceção. (Calvin Mooers apud Sanderson; Croft, 2012, tradução nossa)

Algumas observações podem ser feitas a partir da fala de Mooers. A primeira é a necessidade identificada da utilização de máquinas na recuperação da informação. A segunda reforça a análise feita por Vannevar Bush à forma como a informação era organizada pelas bibliotecas e das dificuldades enfrentadas na recuperação da informação. A RI surge então como o principal braço da CI, sendo considerada como sua atividade mais importante para Saraveric:

“[...] parte dos esforços e recursos da ciência da informação foi, e ainda é, voltado para os problemas de recuperação da informação (RI). A RI não é apenas uma atividade da ciência da informação, mas é a mais importante delas e também onde mais ocorrem as relações interdisciplinares.” (Saraveric, 1995, p. 2)

A recuperação da informação surge sendo desenvolvida em conjunto com a tecnologia, acompanhando suas inovações e criando soluções para os problemas informacionais gerados por ela. Sanderson e Croft 2012 discorrem em seu trabalho sobre os acontecimentos importantes para a área. Em 1952, quando as bibliotecas ainda usavam esquemas de classificação por hierarquia de classes como o Sistema Decimal de Dewey, surge no campo da Biblioteconomia o sistema Uniterm, que basicamente propunha a indexação da informação por uma lista de palavras-chave.

A década de 1960 foi marcada pela transição de questionar se a IR era possível em computadores para a busca de maneiras de melhorar esses sistemas. Gerard Salton foi uma figura chave, contribuindo com conceitos importantes como o modelo vetorial para ranqueamento de documentos e a introdução do feedback de relevância. A partir

de 1970, a RI passa pela formalização de seus modelos, com ênfase em dois principais desenvolvimentos: a combinação de pesos de frequência de termos (tf) de Luhn com a frequência inversa de documentos (idf) de Spärck Jones, e a modelagem probabilística de sistemas de IR. Essas inovações melhoraram significativamente a precisão da recuperação de documentos relevantes, com o modelo vetorial de Salton e o princípio de classificação probabilística de Robertson sendo fundamentais para a pesquisa e desenvolvimento de sistemas de IR nas décadas seguintes.

Os principais acontecimentos na área da RI entre 1980 e 1990, foram o desenvolvimento e refinamento dos esquemas de ponderação tf-idf, a introdução do BM25 como uma função de ranqueamento eficaz, e o avanço do modelo de espaço vetorial com a Indexação Semântica Latente (LSI). Além disso, houve abordagens de linguística computacional e a criação da Text Retrieval Conference (TREC) para lidar com coleções de documentos de grande escala. A recuperação da informação na década de 90 foi impulsionada pelo crescimento da World Wide Web a partir de 1993. O aumento exponencial de sites levou à necessidade de novas técnicas para indexar e recuperar informações, como a análise de links e o uso de textos âncora, que foram fundamentais no desenvolvimento do Google e de outros motores de busca. Esses desenvolvimentos transformaram a IR em um campo mais complexo e dinâmico, adaptando-se às mudanças no ambiente computacional e nas necessidades dos usuários.

Um processo intrínseco à RI é a organização. A organização prévia da informação tem impacto direto na recuperação de informações pertinentes às necessidades informacionais dos usuários. A organização é essencial ao ser humano mesmo nas atividades mais simples, pois estabelece sistemas que facilitam o acesso e a utilização eficiente de recursos. Quando organizamos roupas, alimentos, móveis, dentre outros itens mais complexos, criamos uma ordem que não só melhora a funcionalidade do nosso ambiente, mas também reduz o tempo e o esforço necessários para recuperar e utilizar esses itens. Essa mesma concepção é utilizada pelos Sistemas de Informação (SRIs). Brascher e Carlan destacam que [...] Num sistema de informação, a qualidade obtida na recuperação da informação depende substancialmente dos procedimentos e instrumentos utilizados para organização da informação. (Brascher; Carlan, 2010)

A fim de definir o que é a Organização da Informação (OI), se faz necessária sua diferenciação do conceito de Organização do Conhecimento (OC). Brascher e Café (2008) discutem em seu estudo a diferença entre os conceitos, destacando a existência de uma discordância na definição de Organização da Informação e Organização do Conhecimento. Elas apontam que esses termos são frequentemente utilizados de maneira imprecisa. Para essas autoras, a OI compreende:

[...] a organização de um conjunto de objetos informacionais para arranjá-los sistematicamente em coleções, neste caso, temos a organização da informação em bibliotecas, museus, arquivos, tanto tradicionais quanto eletrônicos. A organização

do conhecimento, por sua vez, visa à construção de modelos de mundo que se constituem em abstrações da realidade.

Esse conceito reforça as concepções expostas no início deste subcapítulo acerca do conhecimento e informação na CI. O tratamento e organização da informação demonstram grande influência sobre a prática e objetivos da CI e da RI.

Os problemas causados pela “explosão informacional” da década de 1950 continuam atuais, como destacou Saraveric (1995), principalmente pelo crescimento da informação digital não organizada e o acesso facilitado desses dados causado pelo advento da rede de computadores. Todas essas informações geradas, de forma contínua pelos mais diversos meios e fontes — em quantidade, variedade, veracidade, valor — como arquivos de textos, códigos, sistemas corporativos, redes sociais, se tornam um grande leque de dados desordenados, que podem ser organizados, analisados e transformados na solução de diversos problemas.

## **2.2 GESTÃO DA INFORMAÇÃO**

A organização da informação desempenha um papel central na administração e geração de conhecimento em diferentes contextos. De acordo com Calazans (2006), a informação é considerada um dos principais ativos organizacionais, o que exige sua gestão e controle para agregar valor aos processos organizacionais e à competitividade. Drucker (1999) define informação como dados organizados dotados de significado, enquanto Miranda (1999) amplia o conceito, descrevendo-a como dados organizados para subsidiar decisões. Já Davenport (1998) aponta que a informação engloba dados, conexões e conhecimento. Historicamente, a humanidade vivenciou quatro revoluções da informação: escrita, livro impresso, imprensa e, atualmente, os canais eletrônicos. Cada uma dessas revoluções influenciou a sociedade e as organizações ao alterar o acesso e a interpretação de dados (Drucker, 1999).

O avanço tecnológico e a globalização tornaram a informação um ativo comparável a recursos financeiros e humanos. Para McGee e Prusak (1994), gerenciar este ativo de forma eficaz é fundamental para as organizações modernas, sendo necessário identificar os fluxos informacionais e compreender as necessidades organizacionais. A organização pode ser definida como o conjunto coordenado de atividades que busca atingir objetivos comuns. Essa definição foi ampliada por Moresi (2001), que considerou a organização um sistema humano planejado que utiliza tecnologia para alcançar metas. Com o passar do tempo, a relação entre organização e informação se tornou mais interdependente, especialmente com a emergência de novas tecnologias e a globalização.

Segundo Calazans (2006), a informação organizacional evoluiu de um elemento burocrático na década de 1950 para um recurso estratégico a partir dos anos 1990. Essa

evolução reflete a crescente importância da informação na tomada de decisões e no planejamento estratégico. Para gerenciar a informação organizacional, Valentim (2002) propõe a classificação em dados estruturados, estruturáveis e não estruturados. Davenport (1998) sugere uma abordagem baseada em fluxos informacionais que inclui fontes impressas, registros e capital intelectual.

Choo (2003) destaca três campos principais no uso organizacional da informação: adaptação às mudanças externas, geração de conhecimento para novos produtos e serviços, e avaliação de alternativas para decisões estratégicas. Essas abordagens demonstram a necessidade de gerenciar os fluxos informacionais para atender às demandas internas e externas das organizações. Drucker (1999) divide a informação organizacional em categorias relacionadas a controle, produtividade, competência e ambiente. Valentim (2002) expande essa visão ao incluir dimensões como informações estratégicas, comerciais, estatísticas e tecnológicas. Moresi (2001) sugere ainda a classificação da informação quanto ao seu nível organizacional e status de importância, priorizando informações críticas e potenciais.

Beal (2004) argumenta que a informação, quando bem gerida, pode promover a troca de ideias, reduzir incertezas e auxiliar na introdução de novos produtos no mercado. Informação estratégica refere-se à capacidade de identificar e usar informações para alcançar vantagem competitiva (McGee; Prusak, 1994). Essa informação inclui dados sobre clientes, concorrentes, aspectos culturais e demográficos, entre outros (Miranda, 1999). A gestão eficiente do fluxo informacional é essencial para garantir que a informação certa esteja disponível no momento certo.

O fluxo informacional, como descrito por Krovi, Chandra e Rayagopalan (2003), deve ser sistematicamente gerido para assegurar a qualidade e relevância da informação. Choo (2003) organiza esse fluxo em etapas que vão desde a identificação de necessidades até o uso efetivo da informação.

## **2.3 DADOS ABERTOS**

Dados abertos são informações que podem ser livremente acessadas, reutilizadas e redistribuídas por qualquer pessoa, respeitando no máximo a atribuição da autoria e a obrigação de compartilhar pela mesma licença. No contexto da Administração Pública, essa prática tem por objetivo principal promover a transparência, a participação cidadã e a prestação de contas, valores fundamentais para o fortalecimento da democracia. Segundo a Open Knowledge Foundation (2019), dados abertos governamentais englobam informações produzidas, coletadas ou custodiadas por autoridades públicas e disponibilizadas de forma acessível, permitindo a sua utilização para diversas finalidades.

A LAI (Lei n.º 12.527/2011) e o Decreto n.º 8.777/2016, que instituiu a Política de Dados Abertos (PDA), são os principais marcos regulatórios brasileiros nesse campo. Ambos reforçam o dever da Administração Pública em disponibilizar informações de interesse cole-



tivo ou geral de maneira transparente e acessível. Essas normas visam assegurar que dados relevantes, como informações financeiras, estatísticas, ambientais e climáticas, estejam acessíveis à sociedade em formatos abertos, estruturados e legíveis por máquinas. A Infraestrutura Nacional de Dados Abertos (INDA) organiza padrões e mecanismos que garantem a interoperabilidade entre sistemas governamentais, possibilitando um compartilhamento mais eficiente dessas informações.

Os benefícios da abertura de dados são amplos e variados. A transparência é um dos principais ganhos, já que a disponibilização de informações claras e confiáveis sobre as atividades do governo permite que os cidadãos compreendam e monitorem as ações públicas. A prática fortalece a prestação de contas (accountability), ao proporcionar mecanismos de supervisão social sobre o desempenho das políticas e serviços públicos. Outro impacto relevante é o estímulo à participação cidadã, que permite que os cidadãos contribuam para a formulação de políticas públicas e soluções colaborativas, utilizando as informações abertas como base.

Internacionalmente, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE, 2019) destaca que os dados abertos possibilitam a melhoria da qualidade dos serviços públicos, promovem a eficácia nos gastos públicos e criam um ambiente de confiança entre o governo e a sociedade. Iniciativas como o portal “Orçamento ao seu Alcance”, que amplia o acesso aos dados financeiros do governo federal, exemplificam como os dados abertos podem ser usados para fomentar a transparência e o controle social.

No Brasil, há exemplos de sucesso na utilização de dados abertos. O município de Recife criou um portal abrangente que organiza informações de diversas pastas administrativas, com temas de interesse público, e disponibiliza dados em formatos acessíveis (Prefeitura do Recife, 2025). Em São Paulo, a Secretaria Municipal de Inovação e Tecnologia lançou a “Vitrine de APIs”, permitindo a conexão entre sistemas públicos e fomentando a criação de serviços baseados em dados (Prefeitura de São Paulo, 2025). Do ponto de vista jurídico, a abertura de dados deve sempre observar a proteção da privacidade e a segurança das informações pessoais. A Lei Geral de Proteção de Dados (Lei n.º 13.709/2018) estabelece critérios para o tratamento de dados pessoais, protegendo os direitos dos cidadãos. Assim, o equilíbrio entre transparência e proteção de dados é essencial para que a prática de abertura de dados seja efetiva e confiável.

Na CI, a qualidade dos dados abertos é avaliada com base em dimensões como adequação, relevância, atualização, facilidade de manipulação e entendimento. Parker et al. (2006) sugerem que essas dimensões são essenciais para garantir que os dados não sejam apenas disponíveis, mas também utilizáveis pela sociedade. Conforme Silva, Monteiro e Reis (2020), o Portal Brasileiro de Dados Abertos atende aos princípios fundamentais de Dados Governamentais Abertos (DGAs), mas apresenta lacunas em dimensões específicas de qualidade, como a facilidade de manipulação e a relevância para usuários diversificados.

A disponibilidade de dados de alta qualidade requer que a Administração Pública implemente práticas de curadoria informacional, com o uso de tecnologias avançadas para a organização, indexação e apresentação de informações. Em ambientes computacionais, técnicas de mineração de dados, aprendizado de máquina e visualização interativa têm sido empregadas para aumentar a acessibilidade e a compreensão dos dados. Assim, a integração dessas tecnologias com portais de dados abertos permite uma análise mais aprofundada (Silva; Monteiro; Reis, 2020).

Do ponto de vista da Computação, a interoperabilidade é um aspecto técnico essencial na disponibilização de dados abertos. Sistemas interoperáveis garantem que diferentes conjuntos de dados possam ser combinados, analisados e reutilizados em novos contextos, aumentando o valor das informações. A arquitetura de interoperabilidade proposta pelos Padrões de Interoperabilidade de Governo Eletrônico (ePING) no Brasil é um exemplo de como padrões técnicos podem ser usados para promover a integração de sistemas governamentais (Silva; Monteiro; Reis, 2020).

Os dados abertos consistem em informações que podem ser livremente usadas, modificadas e compartilhadas por qualquer pessoa, para qualquer finalidade, desde que atendam aos critérios de abertura legal e técnica. De acordo com a definição proposta pela Open Definition (OD), os dados abertos devem ser licenciados sob termos abertos, acessíveis, legíveis por máquinas e disponibilizados em formatos não proprietários (Open Data Watch, 2023).

Os dados abertos são caracterizados por sua disponibilidade em formatos acessíveis, com licenças que permitam uso irrestrito, como as Licenças Creative Commons (CC BY 4.0) e outras de domínio público. Esses dados devem ser disponibilizados de forma que sejam compreensíveis para pessoas e processáveis por máquinas, possibilitando sua reutilização ampla (Open Data Charter, 2019).

No Brasil, a LAI (Lei nº 12.527/2011) é um marco normativo que fortalece a transparência e promove a disponibilização de informações públicas. A criação de portais como o Portal Brasileiro de Dados Abertos exemplifica a aplicação prática desse princípio, centralizando dados de diversas instituições públicas para facilitar o acesso e o uso pela sociedade (Silva, Monteiro e Reis, 2020).

A adoção de dados abertos facilita o acompanhamento de ações governamentais, permitindo maior fiscalização e controle social, estimula o empreendedorismo ao fornecer informações para o desenvolvimento de novas soluções e serviços, reduz custos associados a solicitações de dados e elimina redundâncias no fornecimento de serviços públicos e auxilia na avaliação e no aprimoramento de políticas públicas, possibilitando análises baseadas em dados concretos (Open Data Watch, 2023).

A abertura de dados deve respeitar a privacidade individual e atender às exigências legais, como a Lei Geral de Proteção de Dados (LGPD). Muitos governos enfrentam dificuldades para converter arquivos em formatos abertos e oferecer metadados adequados. A

implementação efetiva exige compromisso político e institucional para garantir continuidade e alinhamento estratégico (Silva, Monteiro e Reis, 2020).

O movimento de dados abertos é sustentado por padrões internacionais, como a Open Data Charter e o Open Government Partnership (OGP). Essas iniciativas promovem princípios fundamentais, incluindo transparência, participação cidadã e inovação tecnológica. Indicadores como o Open Data Barometer (ODB) e o Open Data Inventory (ODIN) avaliam a adoção de práticas de dados abertos e seu impacto nos setores públicos globais. Segundo Silva, Monteiro e Reis (2020), o Portal Brasileiro de Dados Abertos adota a maioria dos princípios internacionais de dados abertos, mas enfrenta desafios em dimensões como atualização, facilidade de manipulação e clareza das informações. Isso indica a necessidade de políticas públicas que priorizem a qualidade dos dados, alinhando-se aos padrões globais.

Os dados abertos não apenas promovem a transparência, mas também têm potencial para transformar a forma como governos, empresas e cidadãos interagem com informações públicas. A disponibilização de dados governamentais em formatos abertos, por meio de plataformas digitais, estabelece um novo paradigma de gestão, baseado na colaboração, inovação e acessibilidade.

A abertura de dados é baseada em duas dimensões fundamentais: a legal e a técnica. A abertura legal refere-se à necessidade de os dados estarem sob licenças que permitam seu uso, redistribuição e modificação. Licenças como Creative Commons (CC BY 4.0, CC0) e as Licenças de Governo Aberto do Reino Unido e Canadá são exemplos amplamente aceitos. A abertura técnica garante que os dados estejam disponíveis em formatos legíveis por máquina, como CSV, XLSX e XML, bem como em formatos abertos que não restrinjam o acesso por questões financeiras ou tecnológicas. Isso assegura a interoperabilidade e facilita o processamento dos dados em diversas ferramentas, incluindo aquelas de código aberto.

Um formato aberto consiste em uma maneira de armazenar dados digitais cuja especificação técnica é pública, sem controle exclusivo de qualquer entidade. Exemplos de formatos abertos incluem CSV, JSON e XML. Em contraste, formatos fechados, como XLS ou PDF, apresentam restrições técnicas ou custos para acesso e reutilização, limitando sua aplicação em contextos de dados abertos.

**Tabela 1 – Formato de dados abertos**

<b>Formato</b>	<b>Exemplo</b>	<b>Uso Ideal</b>
Aberto	CSV, JSON, XML	Análise e integração de dados
Fechado	XLS, PDF	Documentação e leitura humana

Fonte: Feito pelo autor.

Licenças abertas permitem o uso, redistribuição e modificação dos dados sem restrições. Exemplos incluem Creative Commons Attribution (CC BY 4.0) e Open Data Commons Public Domain Dedication and License (PDDL). Essas licenças eliminam barreiras legais para a reutilização dos dados, promovendo maior acessibilidade. A interoperabilidade é um princípio técnico essencial para dados abertos. Ela garante que diferentes conjuntos de dados possam ser integrados e utilizados conjuntamente, independentemente de suas origens, facilitando a criação de soluções inovadoras e análises complexas.

A abertura de dados governamentais permite que cidadãos e instituições monitorem e avaliem a gestão pública, promovendo accountability e combatendo a corrupção. Ferramentas como o “Mapa da Transparência” são exemplos de iniciativas que utilizam dados abertos para informar a população sobre a alocação de recursos públicos (Rio Grande do Sul, 2016).

A reutilização de dados públicos elimina a redundância na coleta e manutenção de informações, reduzindo custos administrativos e simplificando processos de tomada de decisão. A publicação proativa de dados atende às exigências da LAI (Lei nº 12.527/2011), promovendo eficiência e conformidade legal (Brasil, 2011).

**Tabela 2 – Tipos de uso**

<b>Dimensão</b>	<b>Exemplo de Uso</b>
<b>Mobilidade Urbana</b>	Aplicativos de transporte e planejamento de tráfego
<b>Educação</b>	Monitoramento do desempenho de escolas
<b>Saúde</b>	Localização de unidades de saúde e análise epidemiológica

Fonte: Feito pelo autor.

A abertura de dados exige investimentos em infraestrutura tecnológica e capacitação de servidores públicos. Processos como a conversão de dados em formatos abertos podem demandar tempo e recursos significativos (Possamai, 2016).

Embora dados abertos sejam um recurso valioso, a anonimização de dados sensíveis é fundamental para proteger a privacidade dos cidadãos. De acordo com a LGPD, dados anonimizados são aqueles cujos titulares não podem ser identificados, dada a aplicação de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento. Normas claras devem ser adotadas para equilibrar transparência e segurança (World Bank, 2015).

A efetividade dos dados abertos depende do uso ativo por parte da sociedade. Estratégias como hackathons, workshops e consultas públicas podem ajudar a aumentar o engajamento e a reutilização dos dados. Iniciativas de comunicação, como publicações em redes sociais e parcerias com a mídia, são essenciais para ampliar o alcance dos dados abertos. O “Colab.re” é um exemplo de plataforma que conecta cidadãos e governos para resolver problemas urbanos utilizando dados públicos (Colab.Re, 2016).

Hackathons e maratonas de dados são excelentes oportunidades para estimular a inovação e identificar novas aplicações para os dados governamentais. Esses eventos incentivam a participação de desenvolvedores, pesquisadores e cidadãos, promovendo soluções criativas e eficazes. Portais centralizados, como o “Dados RS”, organizam e disponibilizam dados abertos de forma padronizada, facilitando o acesso e a reutilização. Metadados detalhados, como cobertura geográfica e temporal, são essenciais para aprimorar a busca e o entendimento dos dados publicados (Rio Grande do Sul, 2016).

Em dezembro de 2007, um grupo de 30 especialistas reuniu-se em Sebastopol, Califórnia, para elaborar um conjunto de princípios que definem as características fundamentais dos dados abertos governamentais. Esses princípios foram estabelecidos com o objetivo de garantir a transparência, acessibilidade e reutilização eficaz das informações públicas.

**Tabela 3 – Princípios Básicos dos Dados Abertos Governamentais**

<b>Princípio</b>	<b>Descrição</b>
<b>1. Dados Completos</b>	Todos os dados públicos devem ser disponibilizados integralmente, abrangendo os três poderes (Executivo, Legislativo e Judiciário). Não devem ser limitados, salvo por estatutos públicos justificados.
<b>2. Dados Não Discriminatórios</b>	Os dados devem estar acessíveis a todos, sem exigência de cadastro, identificação ou qualquer restrição que limite o acesso.
<b>3. Dados Primários</b>	Devem ser apresentados em estado bruto, tal como coletados da fonte, sem pré-processamento, agregação ou modificação, exceto para proteger dados pessoais.
<b>4. Dados Atuais</b>	Os dados precisam ser publicados regularmente e atualizados com frequência suficiente para manter sua utilidade e relevância.
<b>5. Dados Acessíveis</b>	Devem ser disponibilizados gratuitamente, permitindo o maior alcance e atendendo a diversas finalidades, incluindo uso comercial.
<b>6. Dados Compreensíveis por Máquina</b>	Os dados devem ser estruturados e processáveis por máquina, possibilitando manipulação automatizada com custos baixos.
<b>7. Dados Não Proprietários</b>	Não devem estar sob controle exclusivo de nenhuma entidade. Devem ser publicados em formatos abertos, sobre os quais nenhuma organização detenha exclusividade.
<b>8. Dados Livres de Licenças</b>	Devem estar livres de restrições legais, como direitos autorais ou patentes. Restrições relacionadas à privacidade e segurança são aceitas, se devidamente justificadas.

Fonte: Rio Grande do Sul, 2020.

As licenças abertas são essenciais para garantir o uso, redistribuição e modificação dos dados.

Tabela 4 – Licenças abertas

Licença	Domínio	BY (Atribuição)	SA (Compartilhamento Igual)	Comentários
Creative Commons Zero (CC0)	Conteúdo, Dados	Não	Não	Dados em domínio público, sem restrições.
Open Data Commons Public Domain (PDDL)	Dados	Não	Não	Dados em domínio público, sem restrições.
Creative Commons Attribution 4.0 (CC BY 4.0)	Conteúdo, Dados	Sim	Não	Atribuição exigida para conteúdos e dados.
Open Data Commons Attribution License (ODC-BY)	Dados	Sim	Não	Atribuição exigida para bases de dados.
Creative Commons Attribution Share-Alike 4.0 (CC BY-SA 4.0)	Conteúdo, Dados	Sim	Sim	Exige atribuição e compartilhamento sob os mesmos termos.
Open Data Commons Open Database License (ODbL)	Dados	Sim	Sim	Exige atribuição e compartilhamento sob os mesmos termos.

Fonte: Rio Grande do Sul, 2020.

Os oito princípios fundamentais estabelecem as diretrizes essenciais para a publicação de dados abertos governamentais. Aliados às licenças abertas, esses princípios garantem a reutilização ampla e eficaz das informações públicas, promovendo maior transparência, inovação e engajamento social. As licenças abertas, como as descritas acima, são ferramentas que asseguram o cumprimento desses princípios, permitindo que dados públicos sejam amplamente acessíveis e reutilizáveis por cidadãos, empresas e organizações.

### 2.3.1 Open Government Partnership

A Open Government Partnership, ou Parceria para Governo Aberto, foi criada em 2011 sob a liderança do então presidente dos Estados Unidos, Barack Obama. Este consórcio global nasceu com o objetivo de promover a transparência, a prestação de contas, a participação cidadã e a inovação nos governos. Inicialmente composta por oito países fundadores e nove organizações da sociedade civil, a OGP estabeleceu-se como um fórum único que une governos e sociedade civil para implementar princípios de governo aberto.

Segundo Fraundorfer (2017), a OGP se destaca como um paradigma inovador na promoção de princípios democráticos, ao criar mecanismos participativos e colaborativos entre governos e organizações da sociedade civil. No entanto, a sua eficácia é debatida, especialmente em relação ao uso da OGP por alguns governos como ferramenta de propaganda para encobrir problemas como corrupção e falta de transparência. A OGP opera com base em um modelo de cooperação que exige compromissos concretos e mensuráveis dos países participantes. Esses compromissos são definidos em Planos de Ação Nacional (PAN), elaborados em colaboração com a sociedade civil. Esses planos devem atender a pelo menos um dos quatro pilares principais da OGP:

- 1) Transparência: Divulgação proativa de informações governamentais para facilitar o acesso público.
- 2) Prestação de Contas: Estabelecimento de mecanismos claros para responsabilizar agentes públicos.
- 3) Participação Cidadã: Engajamento efetivo da população na formulação de políticas e decisões governamentais.
- 4) Inovação e Tecnologia: Uso de ferramentas tecnológicas para melhorar a governança e os serviços públicos.

Cada país é avaliado pelo Mecanismo de Relatório Independente (IRM), que verifica o cumprimento dos compromissos assumidos e a qualidade da colaboração entre governo e sociedade civil. A governança da OGP é composta por um comitê diretor dividido entre governos e organizações da sociedade civil. A composição reflete a natureza inclusiva e colaborativa da parceria, com governos como os dos Estados Unidos, Reino Unido e Brasil, e organizações como a International Budget Partnership e a Transparency & Accountability Initiative. Essa estrutura busca garantir um equilíbrio entre as partes, promovendo um diálogo constante e ações conjuntas. Fraundorfer (2017) aponta que a estrutura inovadora da OGP tem potencial para fortalecer processos democráticos e promover governança global participativa. Contudo, estudos sugerem que em alguns países os compromissos assumidos não se traduziram em melhorias reais. Governos podem utilizar a OGP como uma “fachada” para ocultar práticas opacas e corrupção, minando a credibilidade da iniciativa.

### 2.3.2 Open Data Barometer

O Open Data Barometer é um índice analítico desenvolvido pela World Wide Web Foundation em colaboração com a rede Open Data for Development (OD4D), com o objetivo de medir o progresso global em relação à implementação de iniciativas de dados abertos. Na sua quarta edição, o ODB avaliou 115 países, analisando a prontidão, a implementação

e o impacto dos programas de dados abertos, com base nos princípios da Open Data Charter (Brandusescu et al., 2017).

O ODB tem como foco principal monitorar três dimensões essenciais dos dados abertos:

- 1) Preparação: Avalia o nível de prontidão dos países para implementar iniciativas de dados abertos, incluindo aspectos legais, políticos e institucionais.
- 2) Implementação: Examina a abrangência e a qualidade dos programas de dados abertos, observando critérios como licenciamento, acessibilidade, usabilidade técnica e atualidade dos dados.
- 3) Impacto: Mede os efeitos práticos dos dados abertos na promoção da transparência, no fortalecimento de negócios e no engajamento cívico (Brandusescu et al., 2017).

O índice combina diferentes abordagens metodológicas, incluindo análises qualitativas e quantitativas, avaliações técnicas de especialistas e dados contextuais, informações contextuais sobre o desempenho dos países em áreas relacionadas, como governança e economia digital.

Apenas 7% dos conjuntos de dados analisados são verdadeiramente abertos, evidenciando que muitos países ainda enfrentam desafios na implementação de dados abertos em conformidade com os padrões da Open Data Charter. A qualidade e a completude dos dados disponibilizados permanecem inadequadas para atender às necessidades da população, com muitos conjuntos de dados incompletos ou desatualizados.

A falta de vontade política é apontada como um dos principais fatores que impedem o avanço das iniciativas de dados abertos, sobretudo em países de baixa e média renda. Dados críticos para a prestação de contas governamentais, como orçamentos públicos e contratos, permanecem escassos em comparação com dados voltados à inovação econômica, como registros de empresas e estatísticas econômicas (Brandusescu et al., 2017).

Embora os dados abertos possuam um impacto significativo na promoção da inovação econômica, como a criação de startups e a geração de emprego, sua influência em políticas sociais, como educação e saúde, é limitada. Isso reflete a necessidade de priorizar a publicação de dados que promovam equidade social e engajamento cívico, especialmente em comunidades marginalizadas (Brandusescu et al., 2017).

### 2.3.3 Open Data Inventory

O Open Data Inventory é uma ferramenta globalmente reconhecida desenvolvida pela Open Data Watch com o objetivo de avaliar a abertura e a cobertura dos dados



estatísticos produzidos por sistemas nacionais de estatística. O ODIN baseia-se em dois critérios principais para avaliar os dados:

- 1) Cobertura: Avalia a amplitude e profundidade dos dados, abrangendo 21 categorias, que incluem áreas sociais, econômicas e ambientais. Exemplos dessas categorias são saúde, educação, recursos naturais e infraestrutura.
- 2) Abertura: Mede a acessibilidade dos dados com base em aspectos como formatos legíveis por máquina, licenciamento aberto e a facilidade de download.

A metodologia do ODIN é baseada em uma abordagem estruturada que permite comparações consistentes entre países e ao longo do tempo.

**Tabela 5 – Principais Elementos do ODIN**

Elemento	Descrição
Categorias de Dados	Cobrem estatísticas essenciais para o planejamento e monitoramento de políticas públicas.
Critérios de Avaliação	Incluem a existência de dados, formatos acessíveis, periodicidade de atualização e granularidade, como dados subnacionais.
Revisões Periódicas	Os resultados são revisados e ajustados regularmente para manter a relevância e alinhamento com os Objetivos de Desenvolvimento Sustentável (ODS).
	Exemplo: Na edição 2022/23, ajustes foram feitos para incluir exceções em dados subnacionais no cálculo de pontuações.

Fonte: Feito pelo autor.

Os resultados do ODIN são apresentados em rankings globais, que avaliam e comparam os países em termos de suas pontuações de cobertura e abertura. Esses rankings servem como um incentivo para que governos aprimorem seus sistemas estatísticos, identifiquem lacunas e adotem práticas de dados abertos. Muitos países aprimoraram suas práticas de publicação de dados, adotando formatos abertos e aumentando a frequência de atualização. Países em desenvolvimento ainda enfrentam barreiras significativas, como infraestrutura limitada e falta de capacidade técnica. O impacto do ODIN vai além de rankings, influenciando diretamente políticas nacionais e internacionais. O ODIN é mais do que uma ferramenta de medição; ele é um catalisador para a mudança na gestão de dados estatísticos. Ao destacar áreas de melhoria e oferecer diretrizes claras, o ODIN ajuda os governos a adotar práticas baseadas em evidências, melhorar a eficiência administrativa e fortalecer a confiança entre o Estado e seus cidadãos (Open Data Watch, 2023).

## 2.4 DADOS ABERTOS DO INEP

O Plano de Dados Abertos do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira representa uma iniciativa para aprimorar a transparência, acessibilidade e eficiência na divulgação de informações educacionais no Brasil. Implementado para o biênio 2020-2021, o plano é um desdobramento da Política de Dados Abertos do Poder Executivo Federal, regulamentada pelo Decreto nº 8.777/2016. Ele segue diretrizes de governança colaborativa e inovação tecnológica, alinhando-se às normas de proteção de dados pessoais e sigilosos, conforme estabelece a Lei nº 12.527/2011.

O plano visa assegurar boas práticas de disseminação de dados, ampliando a abertura e acessibilidade de informações educacionais ao público. O Inep priorizou a abertura de dados de grande relevância social, baseando-se em demandas da sociedade, como as registradas no Sistema Eletrônico do Serviço de Informações ao Cidadão (e-SIC), e em auditorias realizadas pelo Tribunal de Contas da União (TCU). O e-SIC foi desenvolvido pela Controladoria-Geral da União (CGU) e é “[...] um canal online que permite a qualquer cidadão solicitar informações a órgãos e entidades do Poder Executivo Federal” (e-SIC, 2025).

O plano enfatiza a necessidade de uma gestão descentralizada, em que as unidades organizacionais do INEP assegurem a qualidade, integridade e atualização dos dados publicados. Um modelo de governança foi instituído, com o Comitê de Governança Institucional atuando como patrocinador estratégico, enquanto o Gabinete da Presidência é responsável pela implementação técnica e manutenção do catálogo de dados. Entre as bases de dados priorizadas, destacam-se os microdados do ENEM, os censos da educação básica e superior, e indicadores educacionais. Para garantir a confiabilidade e acessibilidade, o plano prevê a integração contínua desses conjuntos de dados ao Portal Brasileiro de Dados Abertos (Brasil, 2020).

## 2.5 ANÁLISE DA INFORMAÇÃO

A análise da informação é um processo metodológico que envolve a coleta, organização, interpretação e síntese de dados para transformar informações brutas em conhecimento estruturado e útil. Este processo é essencial em diversas áreas do conhecimento, particularmente na pesquisa qualitativa, onde a compreensão dos significados e contextos subjacentes desempenha um papel fundamental na construção de teorias e no avanço do saber científico.

Segundo Santos et al. (2018), a Teoria Fundamentada nos Dados (TFD) exemplifica bem os métodos qualitativos ao promover o desenvolvimento de teorias baseadas nos dados coletados. Originalmente proposta por Glaser e Strauss na década de 1960, a TFD sofreu adaptações que resultaram em três principais vertentes metodológicas: clássica,

straussiana e construtivista, cada uma com abordagens distintas, mas igualmente voltadas para a codificação e análise detalhada dos dados.

Na perspectiva clássica, a análise de dados segue um sistema de codificação em duas etapas: codificação substantiva e codificação teórica. A primeira etapa visa identificar conceitos iniciais nos dados, organizando-os em categorias emergentes, enquanto a segunda busca estabelecer relações teóricas entre essas categorias para formar um modelo explicativo do fenômeno em estudo (Glaser e Strauss, 1967). Essa abordagem destaca-se pela ênfase na neutralidade e objetividade do pesquisador, uma característica típica do positivismo.

Por outro lado, a perspectiva straussiana introduz a codificação axial, um procedimento que conecta categorias e subcategorias de maneira sistemática. Strauss e Corbin (2008) descrevem esse método como um movimento indutivo-dedutivo que permite compreender não apenas os fenômenos, mas também os contextos e condições que os influenciam. Essa vertente incorpora o paradigma da codificação, que organiza os dados em componentes como condições, estratégias e consequências, conferindo maior estrutura à análise.

A abordagem construtivista, desenvolvida por Charmaz, difere significativamente das anteriores ao enfatizar a interação entre pesquisador e participantes na construção do conhecimento. Essa perspectiva utiliza a codificação inicial e focalizada para interpretar os significados atribuídos pelos participantes ao fenômeno estudado, favorecendo uma análise mais contextual e dinâmica (Charmaz, 2009). A ênfase no uso do gerúndio para codificar ações reforça a sensibilidade teórica necessária para captar processos em desenvolvimento.

Independentemente da abordagem adotada, a análise da informação na TFD baseia-se em princípios comuns, como a amostragem teórica, a análise comparativa constante e a elaboração de memorandos. Esses princípios asseguram a continuidade e a profundidade da análise, permitindo que os dados sejam tanto produtos quanto produtores de novos insights durante o processo investigativo (Santos et al., 2018).

A importância da análise da informação transcende a TFD. Em um contexto mais amplo, esse processo é fundamental para transformar dados em conhecimento acionável, informando decisões estratégicas e contribuindo para avanços científicos e tecnológicos. A literatura enfatiza que a escolha do método de análise deve estar alinhada com os objetivos da pesquisa, o referencial teórico e a visão de mundo do pesquisador. Como observado por Foley e Timonen (2015), a adequação metodológica é essencial para garantir a validade e a relevância dos resultados obtidos.

Portanto, a análise da informação é uma prática interdisciplinar que exige rigor metodológico, sensibilidade analítica e um entendimento profundo do contexto em estudo. Como demonstrado pelas diferentes vertentes da TFD, as variações nos métodos de análise refletem as múltiplas formas de compreender e representar a complexidade dos fenômenos sociais, reafirmando o papel central da análise da informação no avanço do conhecimento científico.

O ponto de partida para a análise da informação é a compreensão de que existem dois universos a serem conectados: o dos conteúdos informativos dos documentos e o das necessidades expressas pelos usuários. Para estabelecer essa comunicação, emprega-se a representação da informação, que pode ser realizada por meio de linguagens naturais ou artificiais. A linguagem natural é a forma original e não controlada usada pelos autores e usuários, enquanto a linguagem artificial é desenvolvida com vocabulário e sintaxe controlados para uso em sistemas documentários.

## **2.6 O EXAME NACIONAL DO ENSINO MÉDIO**

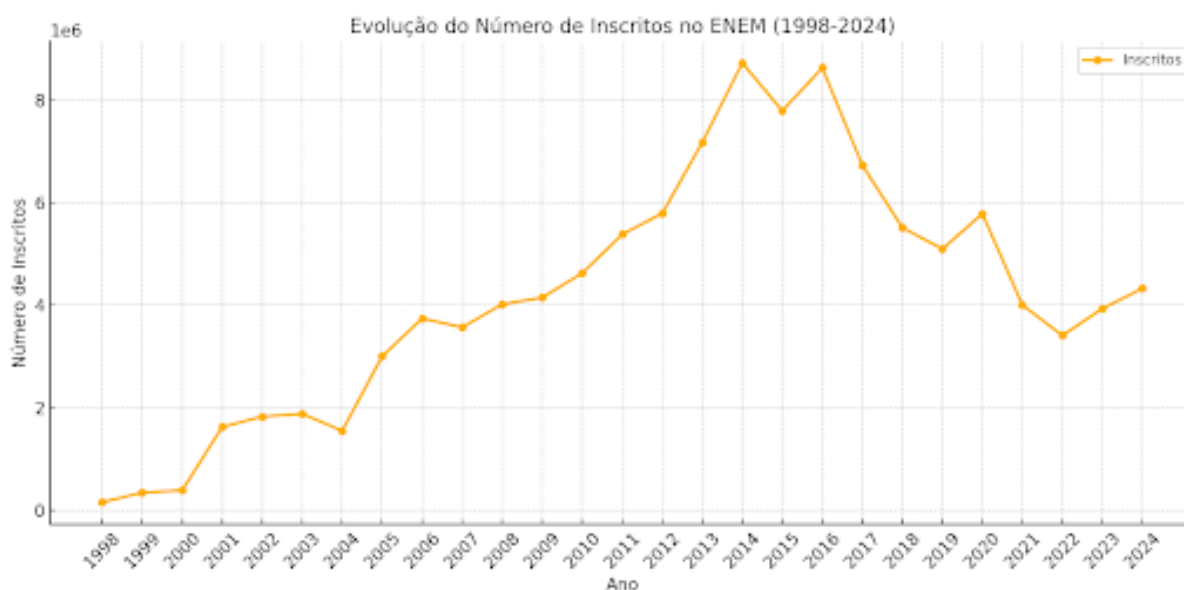
O ENEM foi criado em 1998, pela Portaria MEC nº 438, no governo de Fernando Henrique Cardoso, sob a gestão do então Ministro da Educação Paulo Renato Souza. Desde sua criação, o ENEM foi concebido como um instrumento de avaliação do aprendizado ao final do Ensino Médio, buscando medir competências e habilidades desenvolvidas pelos estudantes ao longo de sua trajetória escolar. Segundo o documento oficial que instituiu o exame, o principal objetivo era “avaliar o desempenho dos alunos, ao término de sua escolaridade básica, obedecendo a uma estrutura de competências associadas aos conteúdos disciplinares” (Brasil, 1998). A tabela 6 sintetiza as mudanças implementadas no exame no decorrer do tempo.

Tabela 6 – Percurso histórico do ENEM

Ano	Principais Mudanças	Impacto
1998	Primeira edição; aplicação em 184 municípios; 157.221 inscritos.	Estabelecimento como avaliação nacional do Ensino Médio.
1999	Aumento de instituições utilizando o exame; aplicação em 162 municípios.	Maior credibilidade e aceitação pelas instituições de ensino.
2000	Inclusão de atendimento especializado; aplicação em 187 municípios.	Inclusão e acessibilidade para pessoas com necessidades especiais.
2001	Inscrição gratuita para concluintes do Ensino Médio; início das inscrições online.	Facilidade e democratização do acesso às inscrições.
2002	Crescimento da abrangência; provas aplicadas em 600 municípios.	Ampliação do alcance e participação de estudantes em todo o Brasil.
2003	Identificação de 'treineiros' no perfil dos participantes.	Mapeamento mais detalhado do perfil dos participantes.
2004	Ligação com o PROUNI; popularização definitiva do exame.	Vinculação do exame a bolsas de estudo; aumento exponencial de inscritos.
2009	Reformulação completa; integração com SISU e PROUNI; ampliação da abrangência.	Mudança de paradigma; acesso ampliado ao Ensino Superior.
2019	Maior taxa de participação da história: 77,23%.	Recorde de participação; consolidação como principal exame educacional.
		Adequação às mudanças

A primeira edição do ENEM, em 1998, contou com 157.221 inscritos e aplicação em 184 municípios. Com um formato simples, composto por 63 questões e uma redação, o exame se propunha a ser uma avaliação diagnóstica das competências e habilidades dos estudantes do Ensino Médio. Apesar de sua abrangência inicial limitada, a prova estabeleceu as bases para sua evolução como um exame nacional (Inep, 2019). A figura 1 apresenta a evolução dos inscritos ao longo do tempo.

**Figura 1 – Evolução do número de inscritos**



Fonte: Inep, 2024

Desde sua criação em 1998, o ENEM apresentou um crescimento expressivo no número de inscritos, especialmente a partir de 2009, com a reformulação do exame para o Novo ENEM. Isso reflete sua transformação em uma ferramenta central para o ingresso no Ensino Superior e sua vinculação a programas como o Programa Universidade para Todos (PROUNI), Fundo de Financiamento Estudantil (FIES) e Sistema de Seleção Unificada (SISU).

Em 1999, o número de instituições de Ensino Superior que utilizaram o ENEM como critério de seleção aumentou para 93, evidenciando a crescente credibilidade do exame. A aplicação foi expandida para 162 municípios, e o criou o Boletim da Escola, oferecendo às instituições dados sobre o desempenho dos alunos (Inep, 2019).

Em 2000, o ENEM incluiu pela primeira vez o atendimento especializado para pessoas com deficiência, marcando um importante avanço em termos de acessibilidade. Além disso, o número de municípios participantes cresceu para 187. A inclusão foi fortalecida em 2001, quando concluintes do Ensino Médio passaram a ter direito à inscrição gratuita,

e o processo de inscrição migrou para um formato online, democratizando ainda mais o acesso (Inep, 2019).

Já em 2002, o exame alcançou 600 municípios, ampliando significativamente sua abrangência. O crescimento da participação foi acompanhado por melhorias na logística e na organização das provas. No ano seguinte, em 2003, o ENEM introduziu um mapeamento mais detalhado do perfil dos participantes, com a identificação de “treineiros” – estudantes que realizavam o exame antes de concluir o Ensino Médio. Essa inovação contribuiu para uma análise mais precisa das condições e motivações dos inscritos (INEP, 2019).

Em 2004, o ENEM foi vinculado ao PROUNI marcando um momento decisivo em sua trajetória. As notas do exame passaram a ser utilizadas como critério para concessão de bolsas integrais e parciais em instituições privadas de Ensino Superior. Esse movimento não apenas popularizou o ENEM, mas também o transformou em um mecanismo de inclusão social, oferecendo oportunidades para estudantes de baixa renda.

Durante esse período, o número de inscritos cresceu exponencialmente, consolidando o exame como uma ferramenta essencial para o acesso ao Ensino Superior. Essa popularização foi acompanhada por um aumento na relevância do ENEM no debate público e nas políticas educacionais. A vinculação com o PROUNI ampliou a função social do ENEM, democratizando o acesso ao Ensino Superior e fortalecendo seu papel na promoção da inclusão educacional (INEP, 2019).

A reformulação de 2009 representou uma transformação profunda na estrutura e nos objetivos do ENEM. Conhecida como Novo ENEM, essa fase trouxe mudanças significativas no formato e ampliou suas funções. A prova passou a ser composta por 180 questões objetivas, organizadas em quatro áreas do conhecimento (Ciências Humanas, Ciências da Natureza, Matemática e Linguagens), além de uma redação. A aplicação foi estendida para dois dias, permitindo maior profundidade na avaliação.

O Novo ENEM também se tornou o principal critério de ingresso em universidades públicas, por meio do SISU. Além disso, as notas passaram a ser utilizadas para acesso ao PROUNI, ao FIES e até em universidades portuguesas com acordos bilaterais. A possibilidade de certificação de jovens e adultos no Ensino Médio foi outro avanço relevante, ampliando as oportunidades educacionais.

A partir de 2010, o ENEM consolidou-se como o maior exame educacional do Brasil, com recordes sucessivos de participação. Em 2019, alcançou 3,9 milhões de participantes presentes, com uma taxa de presença de 77,23%, a maior da história. Essa expansão foi acompanhada pela modernização do exame, incluindo iniciativas como o piloto do ENEM Digital, que visava ampliar o acesso e modernizar a aplicação. Com a implementação do Novo Ensino Médio, o ENEM está passando por uma nova adaptação. A partir de 2024, o exame será totalmente alinhado às diretrizes da Base Nacional Comum Curricular (BNCC), refletindo mudanças no currículo e nos itinerários formativos. Essa transição busca adequar

o ENEM às demandas contemporâneas da educação básica e preparar os estudantes para os desafios do mercado de trabalho e da vida acadêmica (INEP, 2019).

Inicialmente, o ENEM possuía caráter voluntário e era aplicado em um único dia, com uma estrutura composta por 63 questões objetivas e uma redação. A proposta original do exame buscava oferecer uma alternativa de avaliação que transcendesse os modelos tradicionais de provas, com foco em competências e habilidades. Nesse sentido, o MEC destacou que o exame tinha como objetivo “ser um poderoso indutor de mudanças, favorecendo a implantação da reforma curricular do Ensino Médio” (Brasil, 1999). Contudo, sua aplicação inicial limitava-se à autoavaliação e diagnóstico das condições de Ensino Médio no Brasil.

No contexto da década de 1990, a implementação do ENEM estava alinhada às políticas educacionais neoliberais, que enfatizavam a eficiência, a competitividade e a formação para o mercado de trabalho. O exame foi concebido para atender a exigências internacionais de monitoramento da qualidade educacional e controle sobre os sistemas de ensino. Nesse período, “o governo utilizava o ENEM como estratégia de responsabilização das escolas e dos estudantes pelos resultados obtidos”, refletindo uma abordagem de avaliação centralizada no controle e na medição de resultados (Locco, 2005, p. 112).

Em 2009, o exame passou por uma grande reformulação, marcando o início do chamado Novo ENEM. Sob o governo de Luiz Inácio Lula da Silva, e com a gestão do Ministro da Educação Fernando Haddad, o Novo ENEM ampliou suas finalidades e funções. A partir desse momento, a prova deixou de ser apenas uma ferramenta de autoavaliação e passou a ser utilizada como critério para ingresso em instituições de Ensino Superior, por meio do SISU. Foi integrado a programas como o PROUNI e utilizado para a certificação de jovens e adultos no Ensino Médio, substituindo parcialmente o Exame Nacional de Certificação da Educação de Jovens e Adultos (ENCCEJA).

Em 2020, o número de inscritos foi de aproximadamente 5,78 milhões, um número ainda elevado, mas menor do que os anos anteriores. Em 2021, o número caiu significativamente para 4 milhões, indicando um impacto claro da pandemia nas inscrições do exame.

Lima Junior (2015) destaca o papel das provas de ciências da natureza e matemática na reprodução de desigualdades sociais. O autor argumenta que essas áreas apresentam os maiores índices de discriminação em relação à origem social dos estudantes, contribuindo significativamente para o fracasso escolar das classes populares.

De acordo com os dados apresentados por Lima Junior (2015), o desempenho nas provas do ENEM reflete fortemente as diferenças de capital cultural entre as classes sociais, beneficiando estudantes de classes dominantes que possuem maior familiaridade com os conteúdos e formas de avaliação do exame. Desde 2006, o ENEM tem se tornado gradualmente mais discriminatório, com o efeito da origem social sobre o desempenho aumentando ao longo do tempo, especialmente nas provas de ciências da natureza.



Souza Junior (2021) defende que a criação do ENEM foi um marco na educação brasileira, concebido inicialmente como uma ferramenta de diagnóstico e avaliação do Ensino Médio, mas que evoluiu para o principal mecanismo de acesso ao Ensino Superior no país.

(...) o Exame Nacional do Ensino Médio foi criado pelo Ministério da Educação com a publicação da Portaria nº 438, de 1998, com o objetivo de regular a qualidade do Ensino Médio por meio de dados estatísticos gerados a partir da aplicação compulsória da prova na rede nacional de ensino (Souza Junior, 2021, p. 8).

Inspirado em modelos internacionais como o Scholastic Aptitude Test (EUA) e o Baccalaureate (França), o ENEM assumiu um papel central na política educacional, servindo como base para programas como o PROUNI, o FIES e o SiSU. Um dos principais pontos discutidos no artigo é a transformação do ENEM em 2009, que consolidou o exame como um instrumento de avaliação multidimensional. Essa reformulação incluiu a ampliação do número de questões, a introdução de áreas do conhecimento e a adoção de uma abordagem contextualizada e interdisciplinar (Souza Junior, 2021). Essa mudança refletiu um esforço para alinhar o exame às necessidades de um Ensino Médio mais moderno e integrado.

Contudo, a divulgação de rankings baseados nas médias das escolas, por exemplo, gerou distorções pedagógicas, incentivando práticas voltadas exclusivamente para a preparação para o exame. Conforme aponta Souza Junior,

(...) com o intuito de popularizar o ENEM e atrair novos inscritos, em 2005, o INEP/MEC divulgou publicamente, pela primeira vez, as médias por escola e município das provas de 2006. [...] Entretanto, havia um ‘efeito colateral’ danoso do ranqueamento, que só foi corrigido uma década depois (Souza Junior, 2021, p. 7).

Essa prática gerou uma competição entre instituições de ensino, muitas vezes em detrimento de uma formação ampla e voltada para a cidadania. Outro aspecto relevante abordado no artigo é a tentativa de modernização do ENEM com a introdução do formato digital. Embora a digitalização traga potenciais benefícios, como maior segurança e agilidade na correção das provas, sua implementação enfrenta desafios logísticos significativos, especialmente em um país com desigualdades regionais tão marcantes. Segundo o autor,

(...) ao se converter a prova para o formato digital, haveria, sem dúvidas, uma série de benefícios, como a agilidade na correção dos itens de múltiplas escolhas [...]; entretanto, haveria um custo considerável para se equipar os locais de provas com máquinas capazes de ‘rodar’ sem falhas o programa utilizado para se responder à prova digitalmente (SOUZA JUNIOR, 2021, p.12).

O ENEM também se destaca por sua proposta inclusiva, com a oferta de atendimentos especializados para diversos públicos, incluindo pessoas com deficiência, gestantes

e estudantes em classes hospitalares. Essas iniciativas refletem um esforço para tornar o exame mais acessível e alinhado aos princípios de justiça social. Contudo, ainda há limitações e falhas a serem corrigidas.

Lima Junior (2015) destaca que, embora o ENEM tenha se proposto como um instrumento de democratização do acesso ao Ensino Superior, suas características estruturais acabam reforçando as desigualdades educacionais preexistentes, alinhando-se à reprodução da cultura dominante. Nesse contexto, o autor sugere que seria necessário repensar a contextualização e formulação das questões do exame para aproximá-las mais dos saberes e interesses das classes populares.

Outro ponto de crítica é a falta de integração entre o ENEM e as práticas pedagógicas das escolas. Locco (2005) aponta que o modelo de avaliação baseado em competências e habilidades não foi plenamente assimilado pelos docentes, especialmente nas escolas públicas. Segundo a autora, “o ENEM extrapola o âmbito escolar, exigindo competências que muitas vezes não são contempladas nas práticas pedagógicas tradicionais” (Locco, 2005, p. 84). Essa desconexão é reforçada pela ausência de orientações para a adaptação dos currículos escolares às exigências do exame, o que leva muitos professores a recorrerem a materiais e cursinhos externos.

### 3 METODOLOGIA

A metodologia científica é fundamental para garantir a confiabilidade da pesquisa. Ela estabelece um conjunto de práticas sistemáticas que orientam a coleta, análise e interpretação dos dados, assegurando que os resultados obtidos sejam válidos e consistentes.

Os microdados do INEP reúnem um conjunto de informações detalhadas relacionadas às pesquisas, aos exames e às avaliações do Instituto. “Os microdados do ENEM são o menor nível de desagregação de dados recolhidos por meio do exame. Eles atendem a demanda por informações específicas ao disponibilizar as provas, os gabaritos, as informações sobre os itens, as notas e o questionário respondido pelos inscritos no ENEM. (Ministério da Educação, 2024)”.

Nesta pesquisa, foram utilizados os microdados do ENEM 2023, disponibilizados pelo INEP. Esses dados são o objeto de estudo, e a abordagem empregada será a quantitativa. Os microdados fornecidos pelo INEP estão organizados em tabelas, contendo informações coletadas durante a criação e aplicação do exame.

A etapa inicial da análise envolve identificar quais dados são relevantes para compor um panorama de informações. Em seguida, realiza-se um tratamento quantitativo desses dados selecionados, a fim de guiar a criação de gráficos desenvolvidos com a ferramenta Jupyter Notebooks, por *scripts* na linguagem de programação Python e bibliotecas da linguagem como *pandas* e *matplotlib*, específicas para a análise de dados e geração de gráficos.

## 4 ANÁLISE DE DADOS

Conforme o que foi apresentado na Revisão de Literatura, o Exame Nacional do Ensino Médio (ENEM) é uma avaliação anual aplicada no Brasil para medir o desempenho dos estudantes ao final da educação básica e serve como principal porta de entrada para o ensino superior, seja em instituições públicas ou privadas. Os microdados do ENEM são um conjunto de informações detalhadas disponibilizadas pelo INEP, que contém dados anonimizados sobre os participantes do exame, suas respostas, notas e características socioeconômicas, permitindo análises educacionais e estatísticas.

Neste capítulo, será apresentada a análise de parâmetros selecionados dos microdados do ENEM de 2023. Os parâmetros e a devida seleção para esta análise serão apresentados no subcapítulo a seguir.

### 4.1 Conjunto de Dados

É importante esclarecer que as análises foram realizadas considerando os dados fornecidos pelo participante no momento de inscrição, sem considerar sua efetiva realização das provas.

Os microdados do ENEM 2023 contêm informações detalhadas divididas em conjuntos temáticos de variáveis. A seguir, apresenta-se uma descrição organizada desses conjuntos e das variáveis correspondentes.

#### Dados do Participante

- NU\_INSCRICAO
- NU\_ANO
- TP\_FAIXA\_ETARIA
- TP\_SEXO
- TP\_ESTADO\_CIVIL
- TP\_COR\_RACA
- TP\_NACIONALIDADE
- TP\_ST\_CONCLUSAO
- TP\_ANO\_CONCLUIU
- TP\_ESCOLA
- TP\_ENSINO

- IN\_TREINEIRO

#### Dados da Escola

- CO\_MUNICIPIO\_ESC
- NO\_MUNICIPIO\_ESC
- CO\_UF\_ESC
- SG\_UF\_ESC
- TP\_DEPENDECIA\_ADM\_ESC
- TP\_LOCALIZACAO\_ESC
- TP\_SIT\_FUNC\_ESC

#### Dados do Local de Aplicação da Prova

- CO\_MUNICIPIO\_PROVA
- NO\_MUNICIPIO\_PROVA
- CO\_UF\_PROVA
- SG\_UF\_PROVA

#### Dados da Prova Objetiva

- TP\_PRESENCA\_CN
- TP\_PRESENCA\_CH
- TP\_PRESENCA\_LC
- TP\_PRESENCA\_MT
- CO\_PROVA\_CN
- CO\_PROVA\_CH
- CO\_PROVA\_LC
- CO\_PROVA\_MT
- NU\_NOTA\_CN
- NU\_NOTA\_CH

- NU\_NOTA\_LC
- NU\_NOTA\_MT
- TX\_RESPOSTAS\_CN
- TX\_RESPOSTAS\_CH
- TX\_RESPOSTAS\_LC
- TX\_RESPOSTAS\_MT
- TP\_LINGUA
- TX\_GABARITO\_CN
- TX\_GABARITO\_CH
- TX\_GABARITO\_LC
- TX\_GABARITO\_MT

#### Dados da Redação

- TP\_STATUS\_REDACAO
- NU\_NOTA\_COMP1
- NU\_NOTA\_COMP2
- NU\_NOTA\_COMP3
- NU\_NOTA\_COMP4
- NU\_NOTA\_COMP5
- NU\_NOTA\_REDACAO

#### Dados do Questionário Socioeconômico

- Q001-Q025

A escolha das variáveis para a análise de dados foi orientada pelo objetivo de compreender os perfis socioeconômicos e demográficos dos participantes do ENEM de 2023. Essa seleção é fundamental para identificar padrões e particularidades entre os diferentes grupos, contribuindo para um melhor entendimento do contexto educacional e social dos participantes. A seguir serão apresentadas as variáveis, suas respectivas descrições e categorias dos dados.

Tabela 7 – TP\_FAIXA\_ETARIA

Variável	Descrição	Categorias
TP_FAIXA_ETARIA	Faixa etária do participante.	Menor de 17 anos
		17 anos
		18 anos
		19 anos
		20 anos
		21 anos
		22 anos
		23 anos
		24 anos
		25 anos
		Entre 26 e 30 anos
		Entre 31 e 35 anos
		Entre 36 e 40 anos
		Entre 41 e 45 anos
		Entre 46 e 50 anos
		Entre 51 e 55 anos
		Entre 56 e 60 anos
		Entre 61 e 65 anos
		Entre 66 e 70 anos
		Maior de 70 anos

Fonte: Inep, 2023.

Tabela 8 – TP\_SEXO

Variável	Descrição	Categorias
TP_SEXO	Sexo do participante.	Masculino
		Feminino

Fonte: Inep, 2023.

**Tabela 9 – TP\_COR\_RACA**

<b>Variável</b>	<b>Descrição</b>	<b>Categorias</b>
TP_COR_RACA	Cor/raça do participante.	Não declarado
		Branca
		Preta
		Parda
		Amarela
		Indígena
		Não dispõe da informação

Fonte: Inep, 2023.

**Tabela 10 – TP\_ESCOLA**

<b>Variável</b>	<b>Descrição</b>	<b>Categorias</b>
TP_ESCOLA	Tipo de escola do Ensino Médio.	Não Respondeu
		Pública
		Privada

Fonte: Inep, 2023.



Tabela 11 – SG\_UF\_PROVA

Variável	Descrição	Categorias
SG_UF_PROVA	Sigla da Unidade da Federação da aplicação da prova	AC
		AL
		AM
		BA
		CE
		DF
		ES
		GO
		MA
		MT
		MS
		MG
		PA
		PB
		PR
		PE
		PI
		RJ
		RN
		RS
		RO
		RR
		SC
		SP
		SE
		TO

Fonte: Inep, 2023.

Tabela 12 – Q001

Variável	Descrição	Categorias
Q001	Até que série seu pai, ou o homem responsável por você, estudou?	Nunca estudou.
		Não completou a 4ª série/5º ano do Ensino Fundamental.
		Completou a 4ª série/5º ano, mas não completou a 8ª série/9º ano do Ensino Fundamental.
		Completou a 8ª série/9º ano do Ensino Fundamental, mas não completou o Ensino Médio.
		Completou o Ensino Médio, mas não completou a Faculdade.
		Completou a Faculdade, mas não completou a Pós-graduação.
		Completou a Pós-graduação.
		Não sei.

Fonte: Inep, 2023.

Tabela 13 – Q002

Variável	Descrição	Categorias
Q002	Até que série sua mãe, ou a mulher responsável por você, estudou?	Nunca estudou.
		Não completou a 4ª série/5º ano do Ensino Fundamental.
		Completou a 4ª série/5º ano, mas não completou a 8ª série/9º ano do Ensino Fundamental.
		Completou a 8ª série/9º ano do Ensino Fundamental, mas não completou o Ensino Médio.
		Completou o Ensino Médio, mas não completou a Faculdade.
		Completou a Faculdade, mas não completou a Pós-graduação.
		Completou a Pós-graduação.
		Não sei.

Fonte: Inep, 2023.

Tabela 14 – Q006

Variável	Descrição	Categorias
Q006	Qual é a renda mensal de sua família? (Some a sua renda com a dos seus familiares.)	Nenhuma Renda
		Até R\$ 1.320,00
		De R\$ 1.320,01 até R\$ 1.980,00.
		De R\$ 1.980,01 até R\$ 2.640,00.
		De R\$ 2.640,01 até R\$ 3.300,00.
		De R\$ 3.300,01 até R\$ 3.960,00.
		De R\$ 3.960,01 até R\$ 5.280,00.
		De R\$ 5.280,01 até R\$ 6.600,00.
		De R\$ 6.600,01 até R\$ 7.920,00.
		De R\$ 7.920,01 até R\$ 9.240,00.
		De R\$ 9.240,01 até R\$ 10.560,00.
		De R\$ 10.560,01 até R\$ 11.880,00.
		De R\$ 11.880,01 até R\$ 13.200,00.
		De R\$ 13.200,01 até R\$ 15.840,00.
		De R\$ 15.840,01 até R\$19.800,00.
		De R\$ 19.800,01 até R\$ 26.400,00.
		Acima de R\$ 26.400,00.

Fonte: Inep, 2023.

Tabela 15 – Q024

Variável	Descrição	Categorias
Q024	Na sua residência tem computador?	Não.
		Sim, um.
		Sim, dois.
		Sim, três.
		Sim, quatro ou mais.

Fonte: Inep, 2023.

Tabela 16 – Q025

Variável	Descrição	Categorias
Q025	Na sua residência tem acesso à Internet?	Não.
		Sim.

Fonte: Inep, 2023.

## 4.2 Arquitetura da Informação da Análise

O INEP, como já apresentado anteriormente, contém uma série de microdados disponíveis, como o Censo de Educação Superior, Enade, Encceja, Enem, Enem por escola, dentre outros, em formato CSV, acompanhado de seus respectivos dicionários de dados. Os microdados escolhidos são referentes ao Enem de 2023, disponíveis neste endereço: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>.

Após o download dos microdados, foi criado um ambiente virtual para instalação das bibliotecas necessárias com o pacote de instalação pip e execução um script na linguagem de programação Python, acompanhada da biblioteca pandas que é uma ferramenta de análise de dados da linguagem. Utilizando o código abaixo, foi possível extrair os dados CSV, verificar e comparar os valores dos parâmetros selecionados com as informações do dicionário de dados e organizar os dados pertinentes para a análise.

### Código 4.1 – Exemplo de processamento e extração dos microdados utilizando a biblioteca pandas.

```

1 data = pd.read_csv(
2     'MICRODADOS_ENEM_2023.csv',
3     encoding='latin1',
4     sep=';'
5 )
6
7 print(data['TP_SEXO'].head(10))
8
9 # Output
10 0      M
11 1      M
12 2      F
13 3      F
14 4      F
15 5      F
16 6      F
17 7      M
18 8      F
19 9      M
20 Name: TP_SEXO, dtype: object

```

Na sequência, foram utilizadas outras ferramentas da biblioteca pandas para realizar a análise dos dados. Os *scripts* foram transferidos para a plataforma web Jupyter Notebook. O Jupyter Notebook é um ambiente de desenvolvimento interativo que permite a execução

de *scripts* Python e a criação de gráficos. Dentro do *notebook*, foi feita a instalação da biblioteca Matplotlib. Essa é uma biblioteca Python específica para criação de gráficos e visualizações de dados em geral.

Com a análise de dados realizada, foi desenvolvida a criação dos *scripts* para criação dos gráficos. O exemplo completo do *script* Python pode ser visualizada abaixo:

**Código 4.2 – Exemplo de script utilizado para análise e geração de gráficos com as bibliotecas pandas**

```

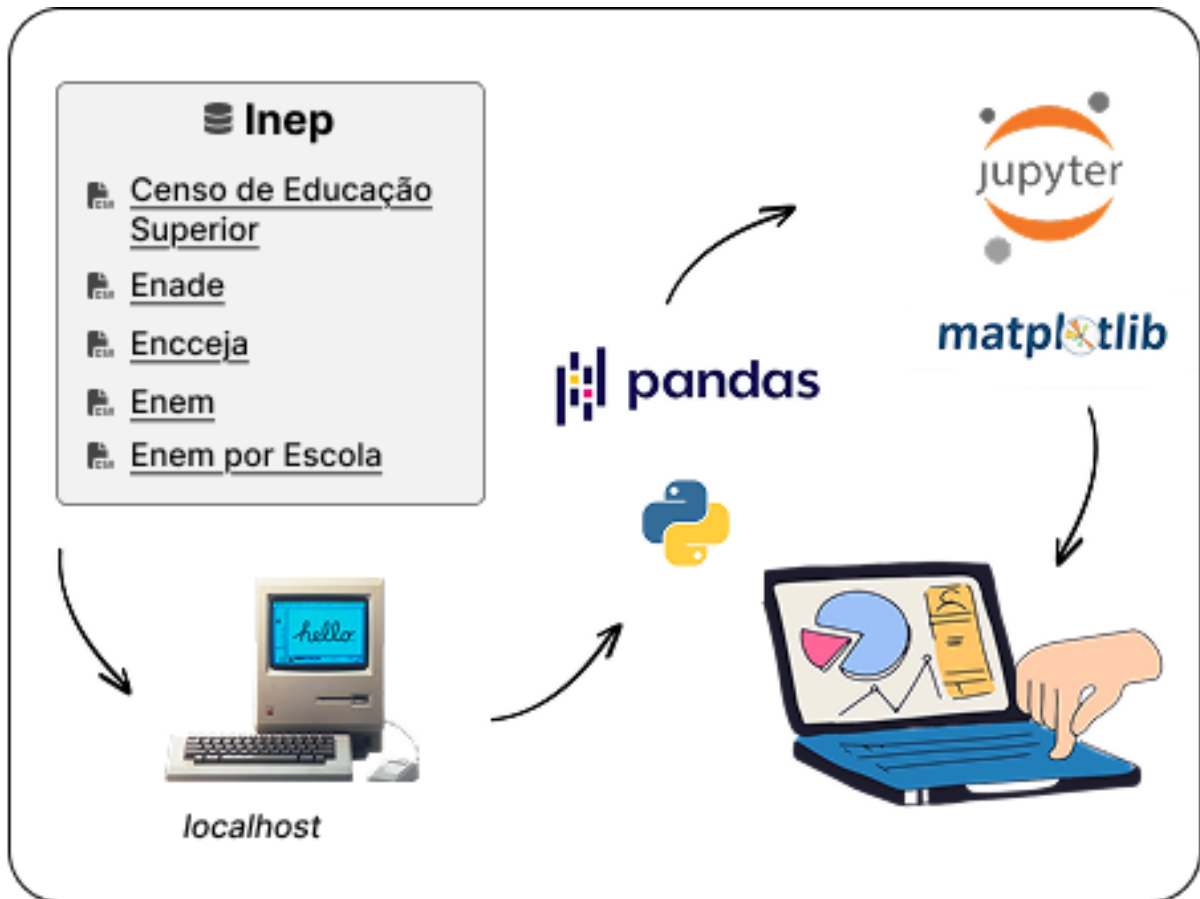
1 # Distribuição de Faixa Etária dos Participantes do ENEM 2023
2
3 age_distribution_map = {
4     1: "<17",
5     2: "17",
6     3: "18",
7     4: "19",
8     5: "20",
9     6: "21",
10    7: "22",
11    8: "23",
12    9: "24",
13    10: "25",
14    11: "26-30",
15    12: "31-35",
16    13: "36-40",
17    14: "41-45",
18    15: "46-50",
19    16: "51-55",
20    17: "56-60",
21    18: "61-65",
22    19: "66-70",
23    20: ">70"
24 }
25
26 data["Faixa Etária"] = data["TP_FAIXA_ETARIA"].map(age_
    distribution_map)
27 age_distribution_order = list(age_distribution_map.values())
28
29 age_distribution = data["Faixa Etária"].value_counts()
30 age_distribution = age_distribution.reindex(age_distribution_
    order, fill_value=0)
31
32 plt.figure(figsize=(12, 8))
33 bars = plt.bar(age_distribution.index, age_distribution.values,
    color="#B5EAEA", edgecolor="black", width=0.6, zorder=2, alpha
    =0.8)
34
35 add_top_bar_labels(bars, 5000)
36
37 plt.title("Distribuição de Faixa Etária", fontsize=16)
38 plt.xlabel("Faixa Etária em Anos", fontsize=12, labelpad=20)
39 plt.ylabel("Número de Inscritos", fontsize=12, labelpad=20)
40 plt.xticks(ha="center", fontsize=10)
41 plt.yticks(fontsize=10)
42 plt.grid(zorder=0)
43 plt.tight_layout()
44 plt.savefig('1.png')

```

```
45 plt.show()
```

A seguir é mostrado o fluxo realizado a fim de alcançar os objetivos descritos:

Figura 2 – Arquitetura da Informação da Análise



Fonte: Feito pelo autor.

### **4.3 Elementos Gráficos**

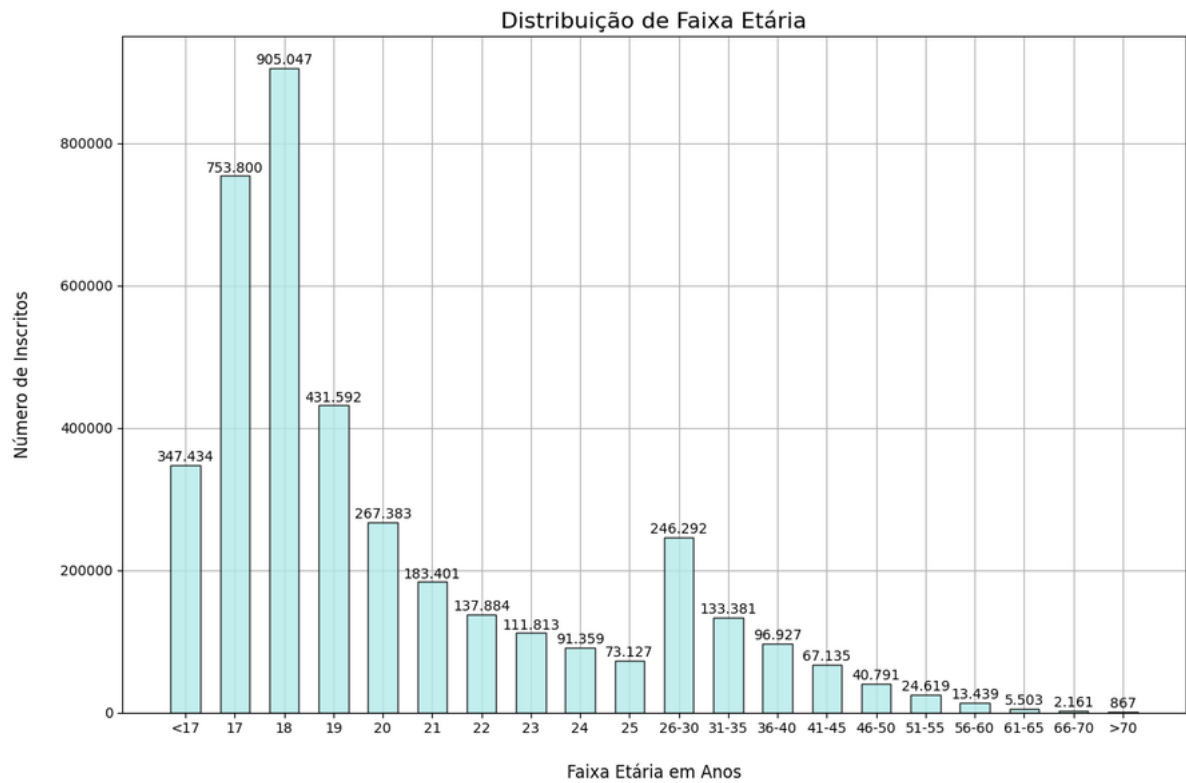
O ENEM 2023 recebeu 3.933.955 milhões de inscrições, portanto essa foi a população considerada na análise. Os microdados obtidos, especialmente os de natureza socioeconômica, incluem respostas que não são de preenchimento obrigatório e carecem de mecanismos de fiscalização para verificar sua autenticidade. Assim, não é possível garantir a veracidade das informações fornecidas, nem assegurar se os estudantes as preencheram com ou sem o auxílio de seus pais, ou responsáveis. No entanto, as perguntas selecionadas para a análise deste estudo são essenciais para delinear um perfil socioeconômico e demográfico básico dos inscritos.



#### 4.3.1 Distribuição de Faixa Etária

A análise da distribuição de faixa etária dos participantes revelou que a maioria se concentra na idade de 18 anos, representando o grupo mais comum entre os inscritos no ENEM, com um total de 905.047 participantes nessa faixa etária. Esse resultado reflete a predominância de jovens recém-egressos ou em fase final do Ensino Médio, que, tradicionalmente, compõem o público-alvo principal do exame. A variável considerada na análise foi a **TP\_FAIXA\_ETARIA**, que representa a faixa etária de cada participante, permitindo identificar a predominância de determinadas idades no universo dos inscritos.

**Figura 3 – Gráfico da distribuição de faixa etária**

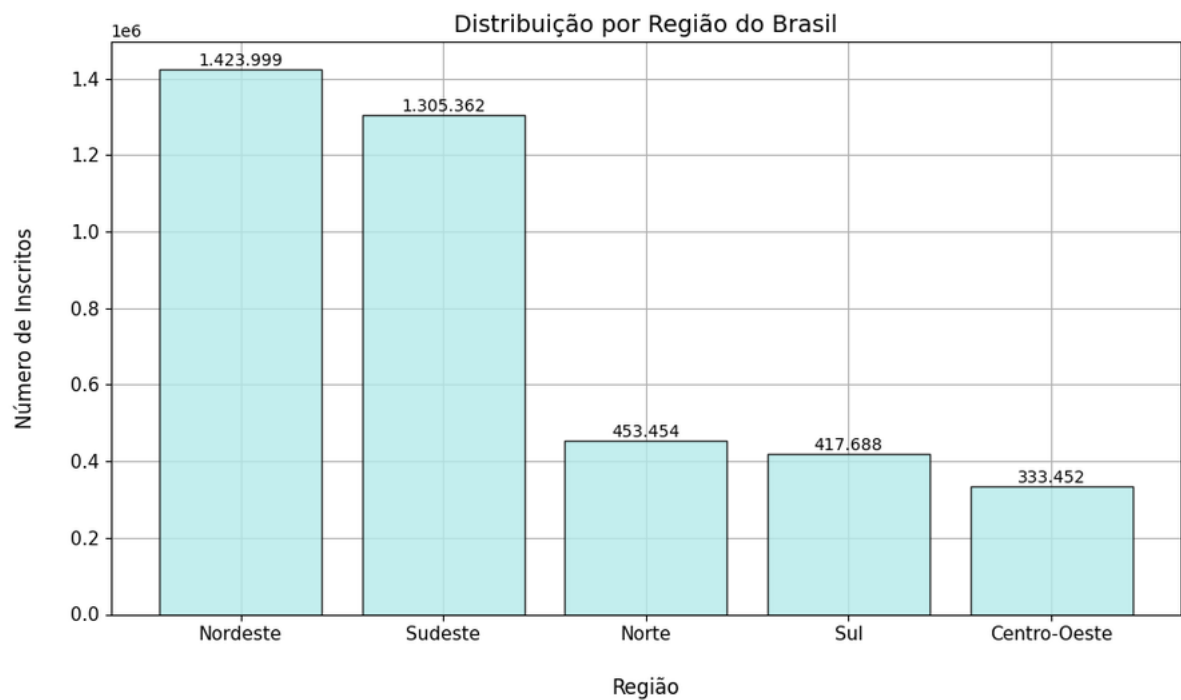


Fonte: Feito pelo autor com base nos dados.

#### 4.3.2 Distribuição por Regiões do Brasil

A análise da distribuição dos inscritos por regiões do Brasil revelou que a maioria dos inscritos vem da região Nordeste, com um total de 1.423.999. Esse dado demonstra a representatividade significativa da região no ENEM, refletindo o acesso e a participação de estudantes de diversas localidades nordestinas no exame. A variável considerada na análise foi a **SG\_UF\_PROVA**, que representa a sigla da Unidade da Federação onde a prova foi aplicada, permitindo identificar a distribuição geográfica dos inscritos.

**Figura 4 – Gráfico da distribuição por regiões do brasil**

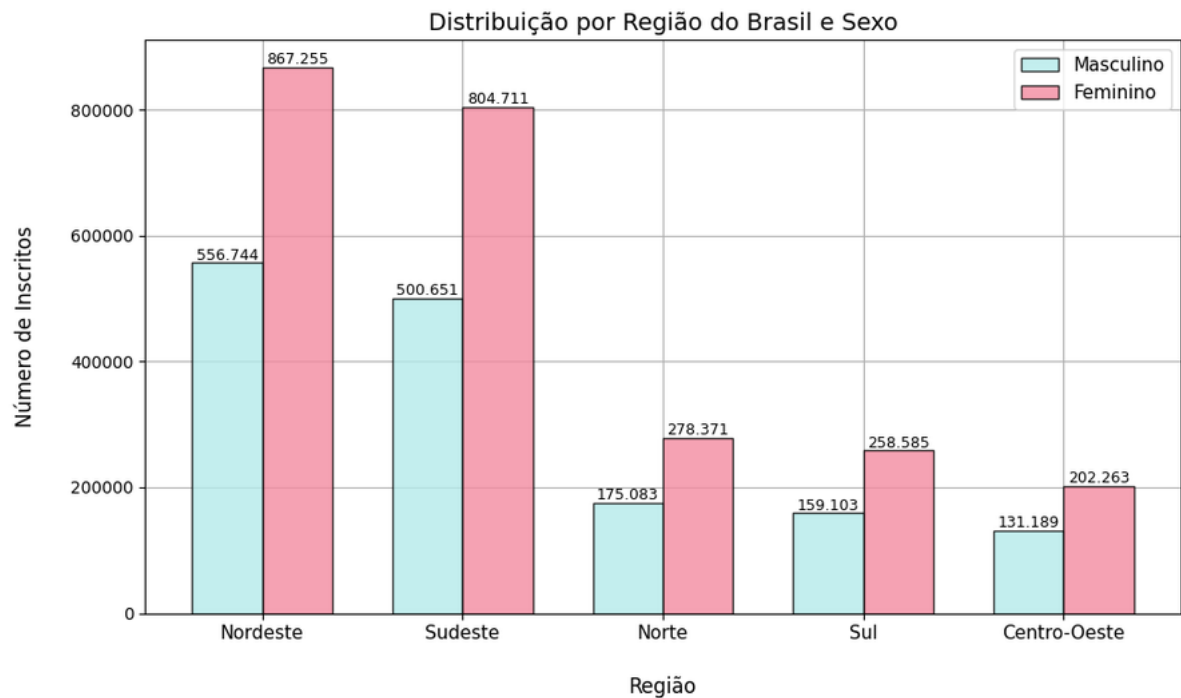


Fonte: Feito pelo autor com base nos dados.

#### 4.3.3 Distribuição por Regiões do Brasil e Sexo

A análise da distribuição dos inscritos por região do Brasil, considerando as variáveis **SG\_UF\_PROVA** e **TP\_SEXO**, mostrou que a maioria dos inscritos do Nordeste são do sexo feminino, totalizando 867.255. O sexo feminino foi predominante em todas as regiões do Brasil, destacando-se como o grupo majoritário no ENEM, independentemente da localidade de aplicação da prova. A variável **SG\_UF\_PROVA** representa a sigla da Unidade da Federação onde a prova foi aplicada, enquanto **TP\_SEXO** identifica o sexo do inscrito, permitindo observar essa tendência de maior participação feminina em todo o território nacional.

**Figura 5 – Gráfico da distribuição por regiões do Brasil e sexo**

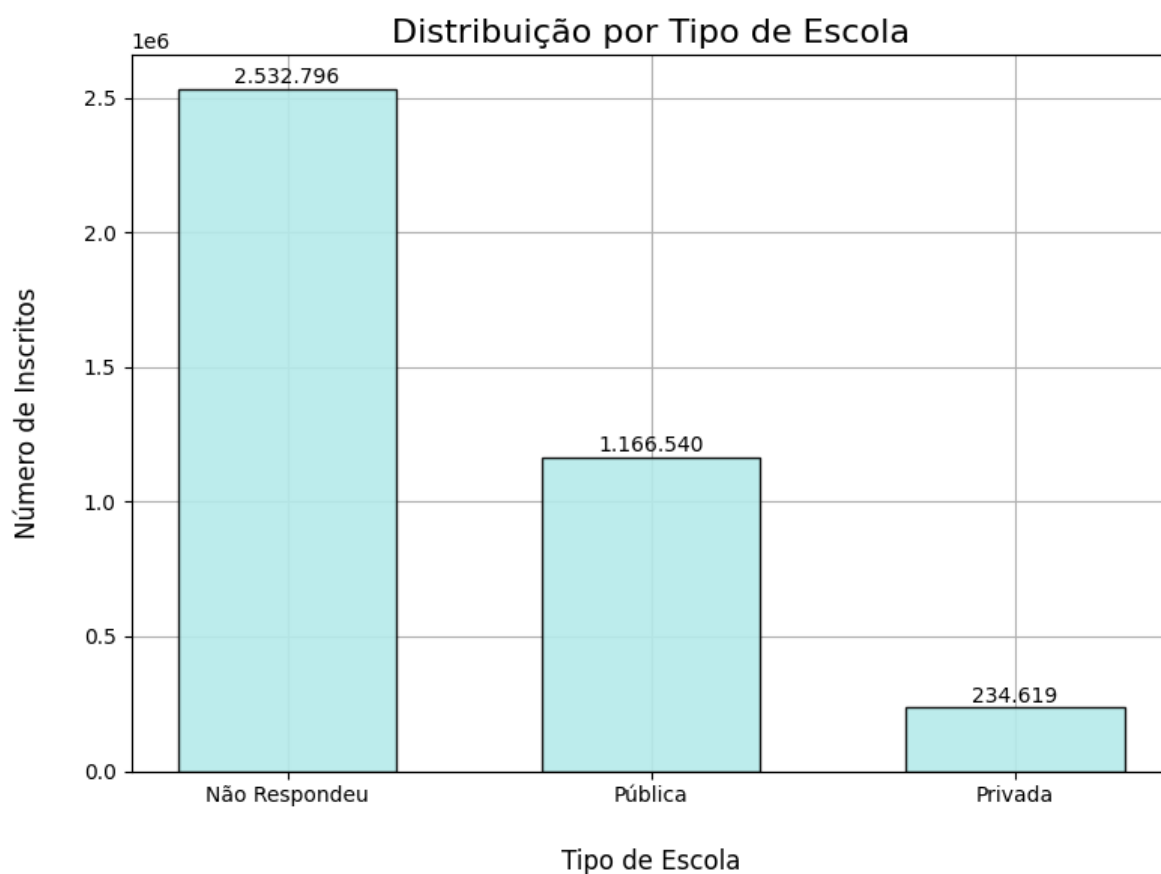


Fonte: Feito pelo autor com base nos dados.

#### 4.3.4 Distribuição por Tipo de Escola do Ensino Médio

A análise da distribuição dos inscritos por tipo de escola, considerando a variável **TP\_ESCOLA**, revelou que a maioria dos inscritos, 2.532.796, não respondeu a essa questão. No entanto, entre os inscritos que forneceram uma resposta, 1.166.540 declararam ter cursado o Ensino Médio em escolas públicas. A variável **TP\_ESCOLA** classifica o tipo de escola onde o inscrito cursou o Ensino Médio, permitindo identificar a predominância de alunos de escolas públicas entre os respondentes. Esse dado pode refletir características socioeconômicas e de acesso à educação, fornecendo informações relevantes para entender o perfil dos inscritos do Enem.

**Figura 6 – Gráfico da distribuição por tipo de escola do Ensino Médio**

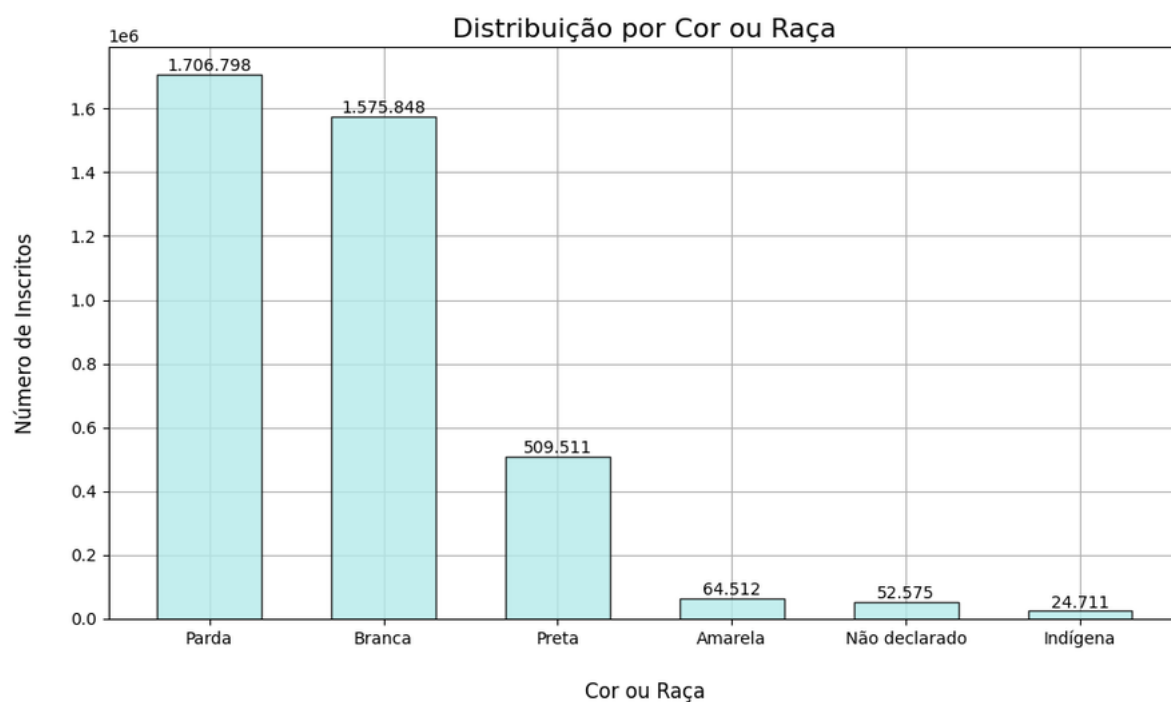


Fonte: Feito pelo autor com base nos dados.

#### 4.3.5 Distribuição por Cor ou Raça

A análise da distribuição de inscritos por cor/raça, utilizando a variável **TP\_COR\_RACA**, indicou que a maioria dos inscritos, 1.706.798, se declararam como pardos. O segundo maior grupo, 1.575.848, corresponde aos inscritos que se declararam como brancos. A variável **TP\_COR\_RACA** classifica os inscritos conforme a sua cor ou raça, oferecendo uma visão sobre a diversidade racial entre os inscritos.

**Figura 7 – Gráfico da distribuição por cor ou raça**

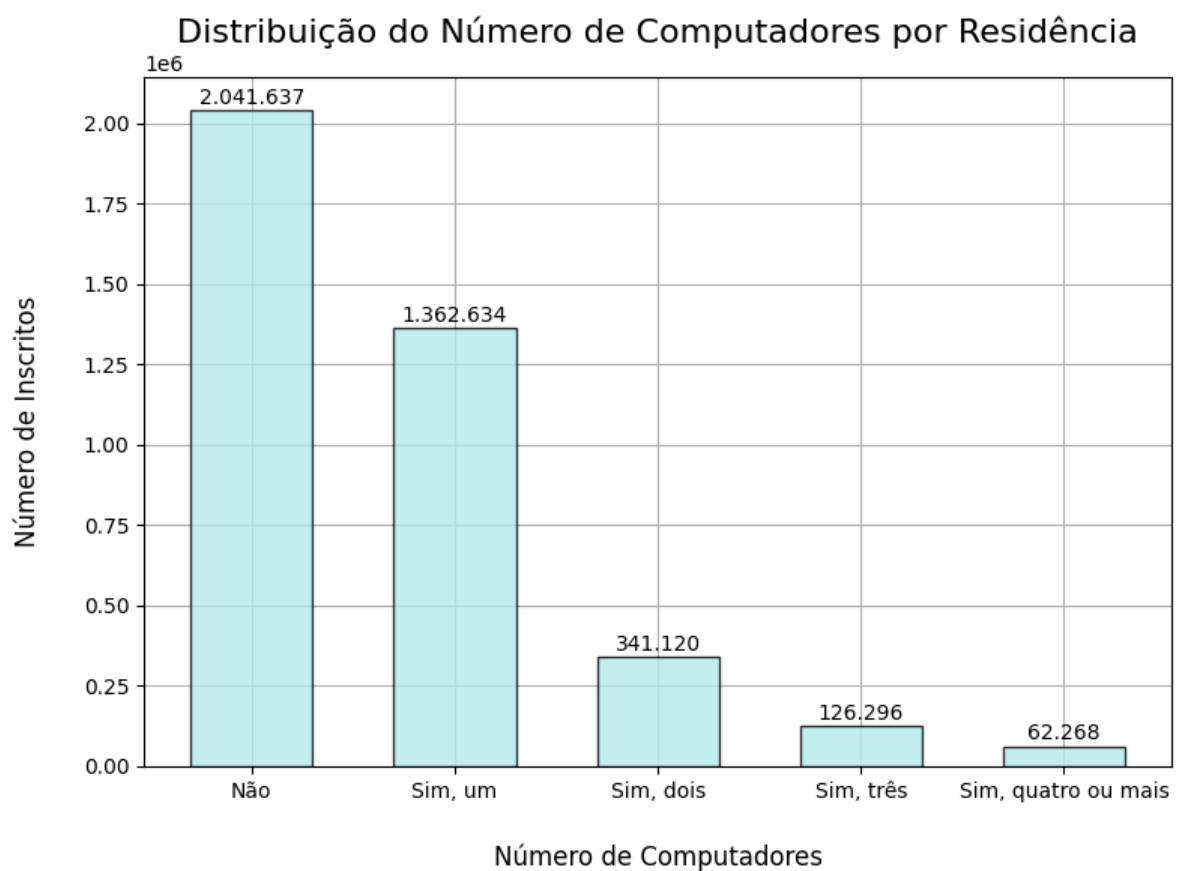


Fonte: Feito pelo autor com base nos dados.

#### 4.3.6 Distribuição do Número de Computadores Por Residência

A distribuição do número de computadores por residência, com base na variável **Q024**, revelou que a maioria dos inscritos, 2.041.637, informou não possuir computador em casa. A variável **Q024** questiona os participantes sobre a disponibilidade de computadores em sua residência, fornecendo dados valiosos sobre o acesso à tecnologia entre os inscritos no ENEM.

**Figura 8 – Gráfico da distribuição do número de computadores por residência**

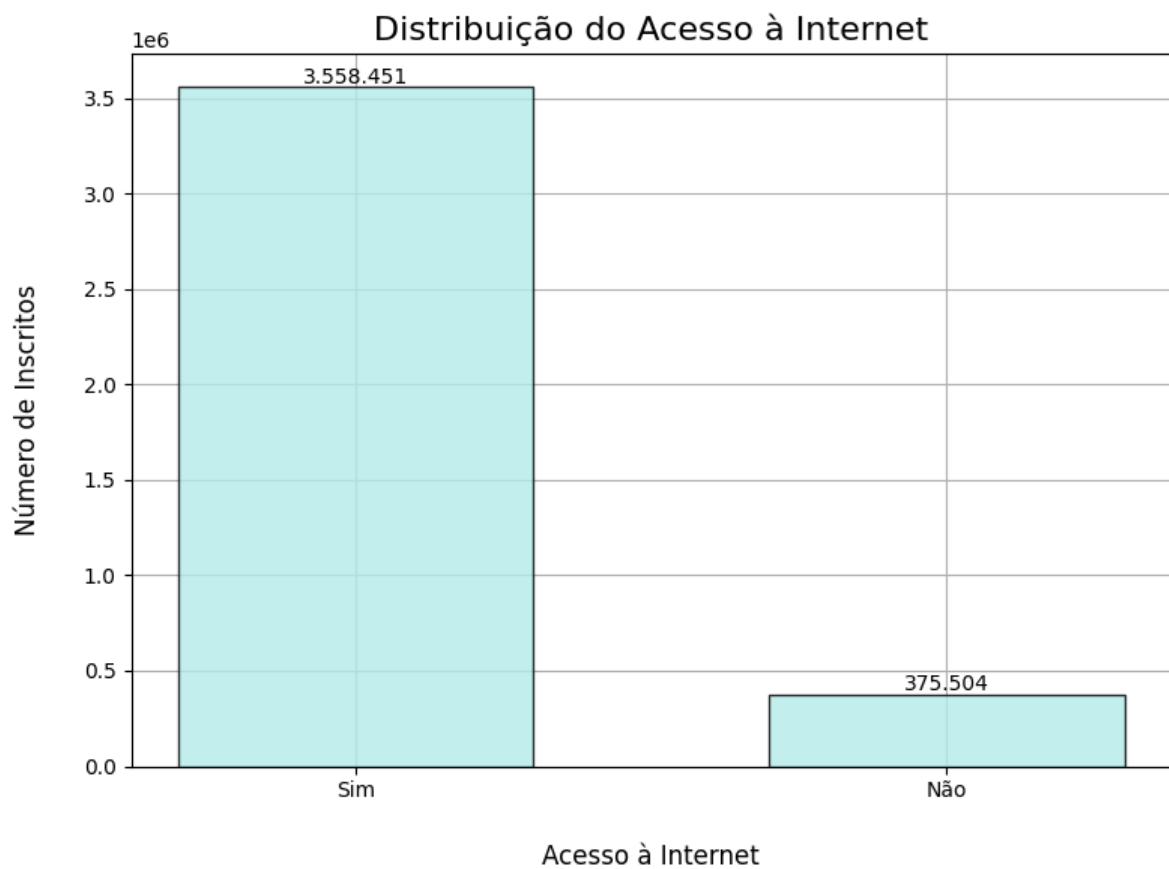


Fonte: Feito pelo autor com base nos dados.

#### 4.3.7 Distribuição do Acesso à Internet

A distribuição do acesso à internet, com base na variável **Q025**, revelou que a maioria dos inscritos, 3.558.451, tem acesso à internet em suas residências. A variável **Q025** questiona se o inscrito possui conexão com a internet em casa, fornecendo dados significativos sobre a disponibilidade de acesso digital entre os participantes do ENEM.

**Figura 9 – Gráfico da distribuição do acesso à Internet**



Fonte: Feito pelo autor com base nos dados.

#### 4.3.8 Distribuição da Escolaridade do Pai

A distribuição da escolaridade do pai, com base na variável **Q001**, indicou que a maioria dos inscritos, 1.114.164, informou que seus pais ou o homem responsável por eles completaram o Ensino Médio, mas não a Faculdade. A variável **Q001** questiona até que série o pai ou responsável do inscrito estudou, oferecendo uma visão sobre o nível educacional dos responsáveis.

**Tabela 17 – Distribuição da Escolaridade do Pai**

<b>Categoria</b>	<b>Descrição</b>	<b>Frequência Absoluta</b>	<b>Frequência Relativa (%)</b>
A	Nunca estudou.	174.712	4,44%
B	Não completou a 4ª série/5º ano do Ensino Fundamental.	700.175	17,80%
C	Completou a 4ª série/5º ano, mas não completou a 8ª série/9º ano do Ensino Fundamental.	513.213	13,05%
D	Completou a 8ª série/9º ano do Ensino Fundamental, mas não completou o Ensino Médio.	438.139	11,14%
E	Completou o Ensino Médio, mas não completou a Faculdade.	1.114.164	28,32%
F	Completou a Faculdade, mas não completou a Pós-graduação.	333.981	8,49%
G	Completou a Pós-graduação.	256.045	6,51%
H	Não sei.	403.526	10,26%

Fonte: Feito pelo autor com base nos dados.



#### 4.3.9 Distribuição da Escolaridade da Mãe

A distribuição da escolaridade da mãe, com base na variável **Q002**, mostrou que a maioria dos inscritos, 1.377.605, informou que suas mães ou a mulher responsável por eles completaram o Ensino Médio, mas não a Faculdade. A variável **Q002** questiona até que série a mãe ou responsável do inscrito estudou, fornecendo informações sobre o nível educacional das mães, o que pode refletir na formação e nas perspectivas educacionais dos inscritos.

**Tabela 18 – Distribuição da Escolaridade da Mãe**

<b>Categoria</b>	<b>Descrição</b>	<b>Frequência Absoluta</b>	<b>Frequência Relativa (%)</b>
A	Nunca estudou.	110.831	2,82%
B	Não completou a 4ª série/5º ano do Ensino Fundamental.	505.525	12,85%
C	Completo a 4ª série/5º ano, mas não completou a 8ª série/9º ano do Ensino Fundamental.	437.022	11,11%
D	Completo a 8ª série/9º ano do Ensino Fundamental, mas não completou o Ensino Médio.	473.854	12,05%
E	Completo o Ensino Médio, mas não completou a Faculdade.	1.377.605	35,02%
F	Completo a Faculdade, mas não completou a Pós-graduação.	455.268	11,57%
G	Completo a Pós-graduação.	441.893	11,23%
H	Não sei.	131.957	3,35%

Fonte: Feito pelo autor com base nos dados.

#### 4.3.10 Distribuição da Renda Familiar

A distribuição da renda familiar, com base na variável **Q006**, indicou que a maioria dos inscritos, 1.245.271 ou 31,65%, possui uma renda mensal de até **R\$ 1.320,00**. A variável **Q006** pergunta sobre a soma da renda do inscrito com a de seus familiares, fornecendo uma visão sobre a situação socioeconômica das famílias dos inscritos. Esse dado é relevante para compreender as condições financeiras dos inscritos e como elas podem influenciar seu acesso à educação e desempenho no Enem.

**Tabela 19 – Distribuição da Renda Familiar**

<b>Categoria</b>	<b>Descrição</b>	<b>Frequência Absoluta</b>	<b>Frequência Relativa (%)</b>
A	Nenhuma Renda	268.053	6,81%
B	Até R\$ 1.320,00	1.245.271	31,65%
C	De R\$ 1.320,01 até R\$ 1.980,00.	650.942	16,55%
D	De R\$ 1.980,01 até R\$ 2.640,00.	437.366	11,12%
E	De R\$ 2.640,01 até R\$ 3.300,00.	293.994	7,47%
F	De R\$ 3.300,01 até R\$ 3.960,00.	171.344	4,36%
G	De R\$ 3.960,01 até R\$ 5.280,00.	261.327	6,64%
H	De R\$ 5.280,01 até R\$ 6.600,00.	139.279	3,54%
I	De R\$ 6.600,01 até R\$ 7.920,00.	85.970	2,19%
J	De R\$ 7.920,01 até R\$ 9.240,00.	75.179	1,91%
K	De R\$ 9.240,01 até R\$ 10.560,00.	59.631	1,52%
L	De R\$ 10.560,01 até R\$ 11.880,00.	39.596	1,01%
M	De R\$ 11.880,01 até R\$ 13.200,00.	33.626	0,85%

<b>Categoria</b>	<b>Descrição</b>	<b>Frequência Absoluta</b>	<b>Frequência Relativa (%)</b>
N	De R\$ 13.200,01 até R\$ 15.840,00.	41.565	1,06%
O	De R\$ 15.840,01 até R\$19.800,00.	41.218	1,05%
P	De R\$ 19.800,01 até R\$ 26.400,00.	38.105	0,97%
Q	Acima de R\$ 26.400,00.	51.489	1,31%

Fonte: Feito pelo autor com base nos dados.

## 5 Conclusão

O presente estudo teve como primeiro objetivo o tratamento dos microdados coletados do site do INEP referentes ao ENEM 2023. Esse processo foi essencial para garantir que apenas as informações mais relevantes ao objetivo geral fossem mantidas, eliminando dados desnecessários e estruturando a base de forma organizada. Dessa maneira, foi possível assegurar maior precisão na análise.

Além disso, a geração de gráficos desempenhou um papel fundamental ao transformar os dados brutos em representações visuais. A visualização gráfica permitiu uma compreensão mais clara das informações, tornando a análise mais acessível e eficiente. Com isso, os gráficos se mostraram ferramentas valiosas para subsidiar decisões, auxiliando tanto pesquisadores quanto gestores educacionais na interpretação dos dados.

Por fim, a análise dos dados sobre os participantes do ENEM revela informações valiosas sobre as características socioeconômicas e demográficas dos participantes, fundamentais para compreender o perfil dos candidatos e as condições em que se encontram. Dessa forma, o estudo reforça a importância da Ciência da Informação na organização e interpretação de grandes volumes de dados, destacando seu papel na tomada de decisões educacionais.

Uma das principais limitações deste estudo diz respeito à disponibilidade e qualidade dos dados fornecidos pelo INEP. Embora os microdados do ENEM sejam uma fonte rica de informações, eles podem conter inconsistências, erros de preenchimento ou lacunas que impactam a precisão das análises. Alguns participantes podem omitir informações socioeconômicas ou preenchê-las de forma incorreta, o que pode levar a distorções nos resultados. Dessa forma, é importante considerar que eventuais imprecisões nos dados podem influenciar as interpretações e os insights extraídos a partir deste estudo.

A partir da distribuição das faixas etárias, observamos que a maioria dos participantes possui 18 anos, refletindo a expectativa do exame como um dos principais meios de acesso ao ensino superior. Além disso, a predominância de inscritos mulheres, especialmente no Nordeste, e a concentração de inscritos provenientes de escolas públicas apontam para uma tendência de maior participação feminina e de estudantes de contextos socioeconômicos mais vulneráveis. Isso reflete a crescente democratização do acesso ao ENEM, mas também evidencia desafios persistentes de equidade na educação brasileira.

Outro ponto importante é a distribuição da cor/raça dos inscritos, onde a maioria se autodeclarou como parda, seguida de brancos. Esses números destacam a diversidade racial do Brasil e a necessidade de ações voltadas à inclusão e à redução das desigualdades enfrentadas por determinados grupos raciais no acesso à educação superior. Também é relevante considerar os dados sobre a escolaridade dos pais. A maioria dos inscritos indicou que suas mães completaram o Ensino Médio, mas não concluíram a faculdade 1.377.605, enquanto o mesmo padrão foi observado para os pais, com 1.114.164 completando o

Ensino Médio sem avançar para o ensino superior. Esses dados sugerem que, apesar dos avanços no acesso à educação básica, há uma barreira significativa para o avanço no ensino superior, refletindo em um ciclo que pode limitar o progresso educacional das famílias.

A renda familiar dos inscritos também revela informações cruciais: a maioria das famílias tem uma renda mensal de até R\$ 1.320,00, reforçando a realidade de um exame que reflete as disparidades socioeconômicas do país. A falta de recursos, tanto financeiros quanto tecnológicos, como a ausência de computadores em muitas residências (2.041.637) e as dificuldades no acesso à internet, impacta diretamente a preparação e o desempenho de muitos candidatos. Isso evidencia a necessidade de políticas públicas voltadas para reduzir essas desigualdades e criar condições mais justas para a participação no exame.

Diante disso, é crucial que o MEC e os órgãos responsáveis pela educação tomem medidas concretas para garantir que os inscritos tenham igualdade de condições para participar e se beneficiar do exame. Algumas ações que podem ser adotadas incluem: a ampliação de programas de inclusão digital, oferecendo acesso à internet e equipamentos para estudantes em situação de vulnerabilidade; o fortalecimento de políticas públicas para reduzir a desigualdade de renda, proporcionando bolsas de estudo ou incentivos para estudantes de famílias de baixa renda; e a implementação de ações de apoio psicológico e educacional, especialmente para as mulheres, alunos pardos, e aqueles provenientes de escolas públicas, garantindo que esses grupos recebam o suporte necessário para competir de forma mais justa. Além disso, seria importante a realização de campanhas de conscientização sobre a importância do ENEM, visando aumentar a participação de jovens de diferentes regiões do Brasil e de diversas condições socioeconômicas.

Por fim, o mapeamento dessas variáveis, além de subsidiar a tomada de decisões políticas do MEC, também pode ser usado para fomentar a criação de políticas públicas voltadas para a melhoria das condições educacionais e sociais dos inscritos, com foco na equidade, inclusão e no rompimento das barreiras históricas enfrentadas por populações em situação de vulnerabilidade.

A Ciência da Informação e a Tecnologia da Informação podem atuar de forma complementar para potencializar a análise e a interpretação de dados educacionais, como os microdados do ENEM. Enquanto a Ciência da Informação se dedica à organização, classificação e extração de conhecimento a partir dos dados, a Tecnologia da Informação fornece as ferramentas e os sistemas necessários para o processamento eficiente dessas informações. No contexto deste estudo, o uso de softwares de análise de dados e visualização gráfica exemplifica essa colaboração, permitindo não apenas a filtragem e estruturação das informações, mas também a geração de *insights* acessíveis para pesquisadores e gestores. Além disso, a aplicação de inteligência artificial e aprendizado de máquina pode aprofundar a identificação de padrões e tendências, auxiliando na formulação de políticas públicas mais assertivas. Dessa forma, a sinergia entre essas áreas contribui significativamente

para a democratização do acesso ao conhecimento e para a construção de soluções mais equitativas no campo educacional.

Profissionais da Ciência da Informação podem se inserir nesta área aplicando seus conhecimentos em organização e curadoria de dados, garantindo que os microdados sejam estruturados de forma acessível e compreensível. Eles também podem atuar na extração de informações relevantes e na criação de soluções de recuperação de dados, otimizando o uso das informações para análises mais precisas e eficientes.

Um possível desdobramento deste estudo seria a realização de uma análise longitudinal do ENEM, comparando os microdados de diferentes edições do exame. Esse tipo de estudo permitiria identificar tendências ao longo dos anos, como mudanças no perfil socioeconômico dos candidatos, evolução no desempenho acadêmico e os impactos de políticas educacionais na democratização do acesso ao ensino superior. Além disso, a análise temporal poderia revelar padrões de desigualdade persistentes e fornecer subsídios mais sólidos para a formulação de estratégias voltadas à redução das disparidades educacionais no Brasil.

## 6 Referências

AGUILLO, C. **Câmbios significativos em el mundo empresarial**. Economia Industrial, n. 330, p. 11-18, 1999.

“APILIB — Vitrine de APIs Da Cidade de São Paulo.” *Sp.gov.br*, 2025, [apilib.prefeitura.sp.gov.br/sto](https://apilib.prefeitura.sp.gov.br/sto). Accessed 23 Feb. 2025.

BARNARD, C. I. **The Functions of the Executive**. Cambridge, Mass: Harvard University Press, 1938.

BEAL, A. **Gestão estratégica da informação: como transformar a informação e a tecnologia da informação em fatores de crescimento e de alto desempenho nas organizações**. São Paulo: Atlas, 2004.

BRANDUSESCU, A.; IGLESIAS, C.; ROBINSON, K.; ALONSO, J. M.; FAGAN, C.; JELLEMA, A.; MANN, D. Open Data Barometer 4th Edition — Global Report. World Wide Web Foundation, 2017.

BRASCHER, Marisa; CAFÉ, Ligia. **Organização da informação ou organização do conhecimento?**. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. Anais. . . São Paulo: ENANCIB, 2008. p. 1-14.

BRASIL. **Decreto n.º 8.777, de 11 de maio de 2016**. Diário Oficial da União, Brasília, DF, 2016.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Política e Plano de Dados Abertos do Inep (Biênio 2020-2021)**. Brasília, 2020.

BRASIL. **Lei de Diretrizes e Bases Nacionais**. (LDB) nº 9.394/96. Brasília, 1996.

BRASIL. **Lei n.º 12.527, de 18 de novembro de 2011**. Diário Oficial da União, Brasília, DF, 2011.

BRASIL. MINISTÉRIO DA EDUCAÇÃO. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). **Censo da Educação Superior - 2021 - Divulgação dos Resultados**. Brasília (DF): Ministério da Educação, 2021.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Escola Que Queremos**. Brasília: INEP-MEC, 2013.

BRASIL. **Portaria MEC Nº 438, de 28 de maio de 1998**. Institui o Exame Nacional do Ensino Médio – ENEM. Brasília, 1998.

BROOKES, B. C. **The foundation of Information Science**. Journal of Information Science, v. 2, Part I, p.125-133, 1980.

BUCKLAND, M. K. **Information as thing**. Journal of the American Society for Information Science, v. 45, n. 5, p. 351-360, 1991.

BUSH, Vannevar (1945). "As We May Think," **Atlantic Monthly 176 (July 1945) pp. 101-108**

CALAZANS, A. T. S. **Conceitos e uso da informação organizacional e informação estratégica**. TransInformação, Campinas, v. 18, n. 1, p. 63-70, jan./abr. 2006.

CHARMAZ, K. **Construção da teoria fundamentada: guia prático para análise qualitativa**. Porto Alegre: Artmed, 2009.

CHOO, C. W. **A organização do conhecimento**. São Paulo: Editora Senac, 2003.

COLAB.RE. **Colab.re - Home**. Recife: Colab.re, 2016. Disponível em: <https://www.colab.re>. Acesso em: 01 fev. 2016.

CRISTÓVAM, José Sérgio da Silva; HAHN, Tatiana Meinhart. **Administração Pública orientada por dados: governo aberto e infraestrutura nacional de dados abertos**. Revista de Direito Administrativo e Gestão Pública, v. 6, n. 1, p. 1-24, 2020.

DAVENPORT, T. **Ecologia da informação: por que só a tecnologia não basta para o sucesso na era da informação**. São Paulo: Futura, 1998.

DINIZ, R. V.; GOERGEN, P. L. **Educação Superior no Brasil: panorama da contemporaneidade**. Avaliação: Revista da Avaliação da Educação Superior (Campinas), v. 24, n. 3, p. 573–593, set. 2019.

DRUCKER, P. **Desafios gerenciais para o século XXI**. São Paulo: Pioneira, 1999.

FRAUNDORFER, Markus. **The Open Government Partnership: Mere Smokescreen or New Paradigm?**. Globalizations, v. 14, n. 4, p. 611-626, 2017. DOI: 10.1080/14747731.2016.1236463.

GENTILI, P. **A exclusão includente na América Latina**. In: ARRAIS, J. (org.), Políticas de Inclusão e Segregação Social. São Paulo: Editora Acadêmica, 2023.

GLASER, B.; STRAUSS, A. **The discovery of grounded theory: strategies for qualitative research**. New Brunswick: Aldine, 1967.

"Hub de Dados Abertos Da Prefeitura Do Recife." *Hub de Dados* -, 6 Feb. 2023, [hubdedados.recife.pe.gov.br/](https://hubdedados.recife.pe.gov.br/). Accessed 23 Feb. 2025.



INEP. **Censo da Educação Superior**. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 1995.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Histórico do ENEM**. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem/historico>. Acesso em: 1 de dez. 2024.

JAMES, Laura. **Defining Open Data**. [S.l.], 03/10/2013. Disponível em: <https://blog.okfn.org/2013/10/03/defining-open-data/>. Acesso em: 09 maio. 2024.

KROVI, R.; CHANDRA, A.; RAJAGOPALAN, B. **Information flow parameters for managing organizational processes**. Communications of the ACM, v. 46, n. 2, p. 77-82, 2003.

LE COADIC, Yves-François. **A Ciência da Informação**. Brasília (DF): Briquet de Lemos/Livros, 2002.

LIMA JUNIOR, Paulo. **Crítica sociológica do Exame Nacional do Ensino Médio: uma análise bourdiana**. In: ENCONTRO NACIONAL DE PESQUISA EM EDUCAÇÃO EM CIÊNCIAS, 10., 2015, Águas de Lindóia. Anais [...]. Águas de Lindóia: ENPEC, 2015. p. 1-8.

LOCCO, L. de A. **Políticas Públicas de Avaliação: O Enem e a escola de Ensino Médio**. Tese de doutorado - PUC-SP. 2005. Disponível em: < [http://www.sapientia.pucsp.br/tde\\_bu sca/arquivo.php?codArquivo=1041](http://www.sapientia.pucsp.br/tde_bu sca/arquivo.php?codArquivo=1041).>

MAIMONE, G. D.; SILVEIRA, N. C. Cognição humana e os Paradigmas da Ciência da Informação. Revista Eletrônica Informação e Cognição, São Paulo, v. 6, n. 1, p. 64, 2007. Disponível em: <https://revistas.marilia.unesp.br/index.php/reic/article/view/748>. Acesso em: 15 maio 2024.

MARIUZZO, P. **Novas cores e contornos na Universidade – o perfil do estudante universitário brasileiro**. Ciência e Cultura, v. 75, n. 1, p. 1–12, 2023.

McGEE, J.; PRUSAK, L. **Gerenciamento estratégico da informação: aumente a competitividade e a eficiência de sua empresa utilizando a informação como ferramenta estratégica**. Rio de Janeiro: Campus, 1994.

MIRANDA, Antônio. **A ciência da informação e a teoria do conhecimento objetivo: Um relacionamento necessário**. In: AQUINO, Mirian de Albuquerque. O Campo da Ciência da Informação: gênese, conexões e especificidade. João Pessoa: Editora Universitária, UFPB. 2002, p. 9-24.

MIRANDA, R. **O uso da informação na formulação de ações estratégicas pelas empresas**. Ciência da Informação, v. 28, n. 3, p. 286-292, 1999.

MOEHLECKE, S. **Políticas de ação afirmativa no Brasil: desafios e possibilidades**. Educação e Pesquisa, v. 30, n. 3, p. 411-430, 2004.

MORESI, E. O contexto organizacional. In: TARAPANOFF, K. (Org.). **Inteligência organizacional e competitiva**. Brasília: UnB, 2001a. p. 59-91.

OECD. **The Path to Becoming a Data-Driven Public Sector**. OECD Digital Government Studies, Paris, 2019.

OPEN DATA WATCH. **Open Data Inventory Biennial Report 2022/23**. Open Data Watch, 2023. Disponível em: <https://www.opendatawatch.com>. Acesso em 13 de novembro de 2024.

OPEN GOVERNMENT WORKING GROUP. **Principles of Open Government Data**. Disponível em: <https://opengovdata.org>. Acesso em: 09 dez. 2024.

OPEN KNOWLEDGE FOUNDATION. **What is Open Data?** 2019. Disponível em: <https://okfn.org/opendata>. Acesso em: 09 dez. 2024.

ORTEGA, C. D. **Relações históricas entre Biblioteconomia, Documentação e Ciência da Informação**. DataGramaZero - Revista de Ciência da Informação, v.5, n.5, out., 2004. Disponível em: <[http://www.dgz.org.br/out04/Art\\_03.htm](http://www.dgz.org.br/out04/Art_03.htm) >. Acesso em: 25 ago. 2024.

PARKER, C. et al. **Measuring the quality of information products**. *Journal of Information Quality*, v. 1, n. 1, 2006.

PINHEIRO, L. V. R.; LOUREIRO, J. M. M. **Traçados e limites da ciência da informação**. Ciência da Informação, v. 24, n. 1, 1995.

POPPER, Karl Raimund. *Quantum theory and the schism in physics*. Vol. 3. Psychology Press, 1992.

POSSAMAI, A. J. **Dados Abertos no Governo Federal brasileiro: desafios de transparência e interoperabilidade**. Tese de Doutorado em Ciência Política – UFRGS, 2016.

RIO GRANDE DO SUL. **Dados RS**. Porto Alegre: Governo do Estado do Rio Grande do Sul, 2016. Disponível em: <http://dados.rs.gov.br>. Acesso em: 03 set. 2016.

RISTOFF, Dimas. **O novo perfil do ensino superior brasileiro: 2001-2012**. Brasília: MEC, SESu, 2012. Disponível em: [http://download.inep.gov.br/educacao\\_superior/censo\\_superior/relatorio\\_analitico\\_censo\\_2012.pdf](http://download.inep.gov.br/educacao_superior/censo_superior/relatorio_analitico_censo_2012.pdf). Acesso em: 05 dez. 2024.

M. SANDERSON and W. B. CROFT, **The History of Information Retrieval Research**, in Proceedings of the IEEE, vol. 100, no. Special Centennial Issue, p. 1444-1451, 13 May 2012, doi: 10.1109/JPROC.2012.2189916.

SANTOS, J. L. G.; CUNHA, K. S.; ADAMY, E. K.; BACKES, M. T. S.; LEITE, J. L.; SOUSA, F. G. M. **Análise de dados: comparação entre as diferentes perspectivas metodológicas da Teoria Fundamentada nos Dados**. Revista da Escola de Enfermagem da USP, v. 52, e03303, 2018. Disponível em: <https://doi.org/10.1590/S1980-220X2017021803303>. Acesso em: 09 dez. 2024.

SARACEVIC, Tefko. **Ciência da Informação: origem evolução e relações**. Perspectivas em Ciência da Informação, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1995.

SCHWARTZMAN, Simon; KNOBEL, Marcelo. **Educação superior no Brasil: O papel das universidades estaduais e municipais**. Revista Brasileira de Educação, São Paulo, v. 12, n. 1, p. 25-40, 1999. Disponível em: <https://www.scielo.br/j/rbedu/a/pYgx6WD>. Acesso em: 05 dez. 2024.

SILVA, Ambrozina de Abreu Pereira; MONTEIRO, Doraliza Auxiliadora Abranches; REIS, Anderson de Oliveira. **Qualidade da Informação dos Dados Governamentais Abertos: análise do Portal de Dados Abertos Brasileiro**. Revista Gestão e Análise, v. 9, n. 1, p. 31-47, 2020.

SOUZA, J. F.; ALMEIDA, P. R. **Ensino Superior e mobilidade interna no Brasil: análise das políticas públicas**. Observatorium: Revista Eletrônica de Geografia, v. 15, 2019, p. 32.

SOUZA, L. P.; PASSOS, L.; FERREIRA, R. A. **Segregação no acesso ao ensino superior no Brasil: perfil dos ingressantes**. Revista de Educação, Ciência e Cultura, v. 25, n. 2, p. 157-172, 2020. DOI: <https://doi.org/10.18316/recc.v25i2.5711>.

STRAUSS, A.; CORBIN, J. **Pesquisa qualitativa: técnicas e procedimentos para o desenvolvimento de teoria fundamentada**. 2. ed. Porto Alegre: Artmed, 2008.

VALENTIM, M. L. P. **Inteligência competitiva em organizações: dado, informação e conhecimento**. Datagramazero - Revista da Ciência da Informação, v. 3, n. 4, p. 1-23, 2002.