



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE QUIXADÁ
CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

LETÍCIA PINHEIRO DE OLIVEIRA

**ANÁLISE DESCRITIVA DE INDICADORES EDUCACIONAIS DA EDUCAÇÃO
BÁSICA NO BRASIL**

QUIXADÁ
2026

LETÍCIA PINHEIRO DE OLIVEIRA

ANÁLISE DESCRITIVA DE INDICADORES EDUCACIONAIS DA EDUCAÇÃO BÁSICA
NO BRASIL

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Sistemas de Informação
do Campus de Quixadá da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Sistemas de Informação.

Orientadora: Profa. Dr^a. Livia Almada
Cruz.

Coorientador: Bel. Paulo Ricardo Gorayeb.

QUIXADÁ

2026

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- O48a Oliveira, Letícia Pinheiro de.
Análise descritiva de indicadores educacionais da educação básica no Brasil / Letícia Pinheiro de Oliveira. – 2026.
65 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Sistemas de Informação, Quixadá, 2026.
Orientação: Profa. Dra. Lívia Almada Cruz.
Coorientação: Prof. Paulo Ricardo Gorayeb.
1. Indicadores Educacionais. 2. Análise Descritiva. 3. Regressão Linear. 4. Visualização de Dados. 5. Dashboard Interativo. I. Título.

CDD 005

LETÍCIA PINHEIRO DE OLIVEIRA

ANÁLISE DESCRITIVA DE INDICADORES EDUCACIONAIS DA EDUCAÇÃO BÁSICA
NO BRASIL

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Sistemas de Informação
do Campus de Quixadá da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Sistemas de Informação.

Aprovada em: 21/01/2026.

BANCA EXAMINADORA

Profa. Dr.^a Lívia Almada Cruz (Orientadora)
Universidade Federal do Ceará (UFC)

Bel. Paulo Ricardo Gorayeb (Coorientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Regis Pires Magalhães
Universidade Federal do Ceará (UFC)

Prof. Dr. David Sena Oliveira
Universidade Federal do Ceará (UFC)

Ao meu avô, cuja presença carinhosa e sábia deixou lembranças que me acompanharam em cada etapa desta caminhada.

Ao meu pai, este trabalho é também o cumprimento da promessa de lhe dar orgulho, e carrega todo o amor e a saudade que sinto.

AGRADECIMENTOS

Agradeço primeiramente à minha família, pelo amor incondicional, pelo apoio em todos os momentos e por acreditarem na minha capacidade mesmo quando eu mesma duvidei.

À família que construí em Quixadá, amigos que se tornaram casa longe de casa, agradeço pela companhia nos dias difíceis, pelas conversas que acalmaram a ansiedade e pelas risadas que tornaram esta caminhada mais leve.

À família que me acolheu, minha namorada e sua família, deixo minha gratidão pelo carinho, pela generosidade e por fazerem com que eu me sentisse parte do lar de vocês desde o início.

Aos meus orientadores, agradeço pela dedicação, pela paciência e pelas contribuições fundamentais para a construção deste trabalho, bem como pelos ensinamentos que levarei para além da vida acadêmica.

RESUMO

Este trabalho teve como objetivo desenvolver um *dashboard* interativo para análise e identificação de tendências em indicadores educacionais brasileiros, utilizando dados históricos disponibilizados pelo INEP. A motivação central foi transformar grandes volumes de dados educacionais em visualizações intuitivas e *insights* acionáveis que pudessem apoiar gestores, pesquisadores e formuladores de políticas públicas na compreensão da evolução da qualidade educacional no Brasil. Foram analisados quatro indicadores principais: Taxa de Distorção Idade-Série (TDI), Taxa de Abandono, Taxa de Aprovação e Taxa de Reprovação, abrangendo séries históricas de 2006 a 2024 (TDI) e 2007 a 2024 (demais indicadores). A metodologia incluiu coleta automatizada de dados via *scripts* Python, pré-processamento e análise exploratória, aplicação de regressão linear para identificação de tendências históricas, e desenvolvimento de um *dashboard* interativo utilizando FastAPI (backend) e React com TypeScript (frontend). Os resultados revelaram melhorias consistentes e estatisticamente significativas em todos os indicadores analisados: redução de 67% no TDI, 69% no abandono, 56% na reprovação e aumento de 15,5% na aprovação. As regiões Norte e Nordeste apresentaram as maiores melhorias relativas, indicando processo de convergência regional. Os valores elevados de R^2 (acima de 0,77) demonstram que as tendências identificadas são robustas e sistemáticas. O *dashboard* desenvolvido está disponível publicamente, permitindo exploração interativa dos dados e facilitando análises comparativas regionais e temporais.

Palavras-chave: indicadores educacionais. análise descritiva. regressão linear. visualização de dados. dashboard interativo. educação básica brasileira.

ABSTRACT

This study aimed to develop an interactive dashboard for analyzing and identifying trends in Brazilian educational indicators, using historical data provided by INEP. The central motivation was to transform large volumes of educational data into intuitive visualizations and actionable insights that could support managers, researchers, and public policy makers in understanding the evolution of educational quality in Brazil. Four main indicators were analyzed: Age-Grade Distortion Rate (TDI), Dropout Rate, Approval Rate, and Failure Rate, covering historical series from 2006 to 2024 (TDI) and 2007 to 2024 (other indicators). The methodology included automated data collection via Python scripts, preprocessing and exploratory analysis, application of linear regression to identify historical trends, and development of an interactive dashboard using FastAPI (backend) and React with TypeScript (frontend). The results revealed consistent and statistically significant improvements in all analyzed indicators: 67% reduction in TDI, 69% in dropout, 56% in failure rates, and 15.5% increase in approval rates. The North and Northeast regions showed the greatest relative improvements, indicating a process of regional convergence. High R^2 values (above 0.77) demonstrate that the identified trends are robust and systematic. The developed dashboard is publicly available, enabling interactive data exploration and facilitating regional and temporal comparative analyses.

Keywords: educational indicators; descriptive analysis; linear regression; data visualization; interactive dashboard; brazilian basic education.

LISTA DE ILUSTRAÇÕES

Figura 1 – Sequência de atividades para execução do projeto	33
Figura 2 – Arquitetura do sistema do <i>dashboard</i> , mostrando a separação entre back-end (FastAPI), front-end (React) e camada de análises.	45
Figura 3 – Interface do <i>dashboard</i> apresentando a evolução nacional do TDI (2006-2024) com linha de tendência e estatísticas descritivas.	49
Figura 4 – Mapa do Brasil com coloração temática representando o TDI por estado. . .	50
Figura 5 – Gráfico de comparação regional mostrando médias e dispersões do TDI. . .	51

LISTA DE TABELAS

Tabela 1 – Análise de tendências do TDI por região geográfica (2006–2024).	49
Tabela 2 – Variação regional da taxa de aprovação (2007–2024).	53
Tabela 3 – Variação regional da taxa de reprovação (2007–2024).	54
Tabela 4 – Variação regional da taxa de abandono (2007–2024).	55

LISTA DE QUADROS

Quadro 1 – Síntese comparativa dos trabalhos relacionados	30
Quadro 2 – Descrição dos indicadores utilizados.	35

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Justificativa	16
1.2	Objetivos	17
1.2.1	<i>Objetivo Geral</i>	17
1.2.2	<i>Objetivos Específicos</i>	17
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Indicadores Educacionais	19
2.2	Mineração de Dados	19
2.3	Conceito de Pipeline de Dados	20
2.4	Visualização de Dados e Dashboards Interativos	20
2.4.1	<i>Visualização de Dados</i>	20
2.4.2	<i>Dashboards Educacionais</i>	21
2.5	Análise Estatística e Regressão Linear	22
2.5.1	<i>Estatística Descritiva</i>	22
2.5.2	<i>Regressão Linear</i>	23
2.5.3	<i>Coefficiente de Determinação (R^2)</i>	23
2.6	Tecnologias para Desenvolvimento do Sistema	24
2.6.1	<i>React</i>	24
2.6.2	<i>Plotly</i>	25
2.6.3	<i>Python</i>	25
3	TRABALHOS RELACIONADOS	26
3.1	<i>From E-budgeting to Smart Budgeting: Exploring the Potential of Artificial Intelligence in Government Decision-Making for Resource Allocation</i>	26
3.2	<i>A Decision Model for Municipal Resources Management</i>	27
3.3	<i>An Application of Machine Learning to Study Utilities Expenses in the Brazilian Navy</i>	28
3.4	<i>Dashboard Interativo para Análise de Indicadores Educacionais: Uma Abordagem Baseada em Dados Abertos</i>	28
3.5	Síntese Comparativa	30
4	METODOLOGIA	32

4.1	Natureza da Pesquisa	32
4.2	Abordagem da Pesquisa	32
4.3	Etapas Metodológicas	33
4.3.1	<i>Seleção e Coleta do Conjunto de Dados</i>	34
4.3.1.1	<i>Fonte e Acesso aos Dados</i>	34
4.3.1.2	<i>Processo de Coleta Automatizada</i>	34
4.3.1.3	<i>Indicadores Educacionais Considerados Inicialmente</i>	35
4.3.1.4	<i>Seleção Final dos Indicadores</i>	36
4.3.1.5	<i>Nível de Agregação dos Dados</i>	36
4.3.1.6	<i>Total de Municípios Analisados</i>	37
4.3.2	<i>Análise Exploratória Inicial e Avaliação de Completude</i>	37
4.3.2.1	<i>Processo de Consolidação dos Indicadores</i>	37
4.3.2.2	<i>Análise de Dados Faltantes</i>	37
4.3.2.3	<i>CrITÉrios de Seleção dos Indicadores</i>	38
4.3.2.4	<i>Estatísticas Descritivas Iniciais</i>	38
4.3.2.5	<i>Validação de Intervalos</i>	39
4.3.2.6	<i>Adição de Variáveis Geográficas</i>	39
4.3.2.7	<i>Agregação Temporal</i>	39
4.3.2.8	<i>Exportação dos Dados Limpos</i>	39
4.3.3	<i>Análise Descritiva Aprofundada dos Indicadores</i>	40
4.3.3.1	<i>Análise da Taxa de Distorção Idade-Série (TDI)</i>	40
4.3.3.1.1	<i>Cálculo de Métricas Estatísticas</i>	40
4.3.3.1.2	<i>Análise Regional de Tendências</i>	41
4.3.3.2	<i>Análise dos Indicadores de Fluxo Escolar (Abandono, Aprovação e Reprovação)</i> 41	
4.3.4	<i>Cálculo de Métricas Estatísticas e Análise de Tendências</i>	42
4.3.4.1	<i>Regressão Linear para Identificação de Tendências</i>	42
4.3.4.2	<i>Coefficiente de Determinação (R^2)</i>	42
4.3.4.2.1	<i>Interpretação Descritiva do R^2</i>	43
4.3.4.3	<i>Agregação de Resultados</i>	43
4.3.5	<i>Desenvolvimento de Visualizações</i>	43
4.3.5.1	<i>Gráficos de Evolução Temporal com Tendência</i>	43
4.3.5.2	<i>Gráficos de Comparação</i>	44

4.3.5.3	<i>Mapas Geográficos Interativos</i>	44
4.3.6	<i>Desenvolvimento do Dashboard Interativo</i>	44
4.3.6.1	<i>Arquitetura do Sistema</i>	45
4.3.6.1.1	Back-end (Python com FastAPI)	45
4.3.6.1.2	Front-end (React com TypeScript)	46
4.3.6.2	<i>Funcionalidades Implementadas</i>	46
4.3.6.3	<i>Disponibilidade e Acesso</i>	46
5	RESULTADOS	48
5.1	Análise da Taxa de Distorção Idade-Série (TDI)	48
5.1.1	<i>Resultados Nacionais</i>	48
5.1.2	<i>Análise Regional</i>	49
5.1.2.1	<i>Nordeste e Norte – Maiores Melhorias</i>	50
5.1.2.2	<i>Sul, Centro-Oeste e Sudeste – Melhorias Moderadas</i>	51
5.1.2.3	<i>Consistência das Tendências</i>	51
5.2	Análise das Taxas de Rendimento Escolar	52
5.2.1	<i>Taxa de Aprovação</i>	52
5.2.1.1	<i>Resultados Nacionais</i>	52
5.2.1.2	<i>Análise Regional</i>	52
5.2.2	<i>Taxa de Reprovação</i>	53
5.2.2.1	<i>Resultados Nacionais</i>	53
5.2.2.2	<i>Análise Regional</i>	53
5.2.3	<i>Taxa de Abandono</i>	54
5.2.3.1	<i>Resultados Nacionais</i>	54
5.2.3.2	<i>Análise Regional</i>	55
5.3	Síntese dos Principais Insights	55
5.3.1	<i>Tendências Positivas Generalizadas</i>	56
5.3.2	<i>Convergência Regional</i>	56
5.3.3	<i>Padrões Identificados nos Indicadores de Fluxo Escolar</i>	56
5.3.4	<i>Consistência das Tendências</i>	57
5.3.5	<i>Desafios Persistentes</i>	57
6	CONCLUSÃO	58
6.1	Principais Resultados	58

6.1.1	<i>Tendências Positivas Generalizadas</i>	58
6.1.1.1	<i>Consistência das Tendências</i>	58
6.1.2	<i>Contribuições do Trabalho</i>	58
6.1.2.1	<i>Contribuição Prática</i>	59
6.1.2.2	<i>Contribuição Metodológica</i>	59
6.1.2.3	<i>Contribuição Teórica</i>	59
6.1.3	<i>Limitações do Estudo</i>	59
6.1.4	<i>Trabalhos Futuros</i>	60
6.2	Considerações Finais	60
	REFERÊNCIAS	62

1 INTRODUÇÃO

A era digital transformou profundamente a forma como dados são coletados, armazenados e utilizados para a tomada de decisões estratégicas. No contexto educacional brasileiro, a disponibilização de dados abertos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) representa um avanço significativo para a transparência e o monitoramento da qualidade da educação básica (INEP, 2024). O acesso público a indicadores educacionais permite que gestores, pesquisadores, educadores e a sociedade civil compreendam melhor os desafios enfrentados pelo sistema educacional e identifiquem oportunidades de melhoria na qualidade do ensino, na equidade de acesso e na efetividade das políticas públicas educacionais (Observatório de Educação, 2021).

A Lei de Acesso à Informação (Lei nº 12.527/2011) (BRASIL, 2011) estabeleceu como dever dos órgãos públicos promover a divulgação de informações de interesse coletivo, impulsionando iniciativas de dados abertos governamentais no Brasil (Observatório de Educação, 2021). No campo educacional, diversas plataformas reúnem dados sobre trajetória escolar, contexto socioeconômico e aprendizagem: o Painel Educacional do INEP (INEP, 2024) apresenta indicadores consolidados, o portal QEDu (QEDu, 2024) oferece visualizações interativas de dados educacionais municipais e estaduais e o Observatório do PNE (Observatório de Educação, 2016) monitora o cumprimento das metas do Plano Nacional de Educação.

Embora essas plataformas representem avanços importantes na democratização do acesso à informação educacional, algumas apresentam visualizações estáticas ou filtros limitados que dificultam análises comparativas entre regiões e a observação de tendências ao longo de períodos extensos. Nesse contexto, *dashboards* interativos desenvolvidos com foco em análises regionalizadas e comparações temporais podem complementar as ferramentas existentes, possibilitando que gestores educacionais explorem tendências históricas por meio de visualizações dinâmicas, identifiquem disparidades regionais e realizem comparações detalhadas entre diferentes níveis geográficos de forma intuitiva (Observatório de Educação, 2016; Cortina, 2025).

Entre os principais indicadores educacionais no Brasil, destacam-se as taxas de aprovação, reprovação e abandono escolar, que compõem o cálculo do Índice de Desenvolvimento da Educação Básica (IDEB) (INEP, 2024). O IDEB, criado em 2007 pelo Ministério da Educação, combina o desempenho dos alunos em avaliações padronizadas como o SAEB (Sistema Nacional de Avaliação da Educação Básica) com indicadores de fluxo escolar, refletindo

tanto a aprendizagem quanto a capacidade das escolas em promover a progressão regular dos estudantes (Fonseca; Namen, 2016). A inclusão das taxas de reprovação e abandono no cálculo do IDEB demonstra a qualidade do ensino, a eficiência e o impacto social das instituições educacionais.

A análise longitudinal desses indicadores, observando seu comportamento ao longo de séries históricas, pode revelar tendências importantes sobre a evolução do sistema educacional brasileiro. Dados do INEP apontam que as taxas de abandono, aprovação e reprovação variam significativamente entre regiões, estados e municípios, refletindo desigualdades socioeconômicas e diferenças nas políticas educacionais locais (INEP, 2024; Jeduca, 2023). Compreender como essas métricas se comportam ao longo do tempo e identificar padrões regionais é essencial para direcionar recursos e intervenções de forma mais eficaz.

Além dos indicadores de fluxo escolar, o acompanhamento da Taxa de Distorção Idade-Série (TDI) ao longo do tempo fornece uma visão abrangente sobre a evolução da qualidade educacional. Técnicas de regressão linear permitem identificar relações entre variáveis e analisar tendências baseadas em dados históricos, fornecendo subsídios para a gestão educacional (Revisa, 2025). O coeficiente de determinação (R^2) possibilita avaliar o grau de ajuste de modelos lineares aos dados observados, indicando a força das tendências identificadas (Revisa, 2025).

A análise preliminar dos dados disponibilizados pelo INEP revelou que, dos 14 indicadores educacionais inicialmente considerados, apenas quatro — Taxa de Distorção Idade-Série (TDI), taxas de Abandono, Aprovação e Reprovação — apresentavam séries históricas suficientemente completas para viabilizar análises longitudinais confiáveis. Essa constatação direcionou o escopo do trabalho para uma abordagem descritiva e exploratória, focada na análise de tendências históricas e na identificação de padrões temporais e regionais nos dados disponíveis. A escolha pelo desenvolvimento de um *dashboard* interativo justifica-se pela necessidade de criar uma ferramenta que permitisse explorar visualmente os dados já tratados, facilitando análises comparativas e a identificação de *insights* relevantes para gestores educacionais. O *dashboard* possibilita que usuários realizem suas próprias análises exploratórias de forma intuitiva, sem necessidade de conhecimentos técnicos avançados (Cortina, 2025).

Este trabalho propõe a criação de um *dashboard* interativo para análise e identificação de tendências em indicadores educacionais brasileiros, utilizando dados históricos. A arquitetura do sistema é dividida: o back-end, desenvolvido em Python, é responsável pelo processamento dos dados educacionais (leitura de CSV, transformação e cálculo de métricas); o front-end,

construído em React, consome esses dados via API e os apresenta em visualizações dinâmicas, utilizando a biblioteca Plotly.

Essa separação de responsabilidades assegura que a camada de análise de dados seja especializada e o React se concentre na interface e na experiência do usuário. Como resultado, o sistema ganha em facilidade de manutenção, permite a reutilização de componentes e favorece futuras expansões. A funcionalidade principal é permitir a exploração dinâmica dos dados, o que facilita a rápida identificação de padrões, sejam eles temporais ou regionais.

O *dashboard* desenvolvido incorpora análises de regressão linear aplicadas especificamente para identificação de tendências temporais nos quatro indicadores educacionais selecionados: Taxa de Distorção Idade-Série (TDI), taxas de abandono, aprovação e reprovação. Para cada município e indicador, foi ajustado um modelo de regressão linear simples onde a variável independente é o tempo (ano) e a variável dependente é o valor do indicador, permitindo quantificar se o indicador apresentou tendência crescente, decrescente ou estável ao longo da série histórica (Montgomery *et al.*, 2012). O coeficiente angular da reta de regressão indica a magnitude da mudança anual, enquanto o coeficiente de determinação (R^2) mede a consistência dessa tendência, indicando o grau de ajuste do modelo linear aos dados observados. Os dados foram agrupados por código de município, permitindo análises regionalizadas que consideram as especificidades de estados e regiões do país.

A análise criteriosa dos dados educacionais por meio de visualizações interativas pode auxiliar gestores públicos a direcionar melhor os recursos e a implementar políticas educacionais mais efetivas (Observatório de Educação, 2016; Observatório de Educação, 2021). Ao buscar contribuir para a democratização do acesso a *insights* derivados de dados abertos, este trabalho visa apoiar uma gestão educacional mais transparente, participativa e orientada por evidências, promovendo melhorias na qualidade da educação básica brasileira.

1.1 Justificativa

A crescente disponibilidade de dados educacionais abertos no Brasil cria oportunidades para análises mais profundas sobre o desempenho do sistema educacional (INEP, 2024; Observatório de Educação, 2021). No entanto, a complexidade e o volume desses dados dificultam sua interpretação por parte de gestores e formuladores de políticas públicas que não possuem conhecimentos técnicos avançados em análise de dados (Observatório de Educação, 2016; Observatório de Educação, 2021).

Desenvolver um *dashboard* interativo que consolida indicadores educacionais históricos e apresenta visualizações intuitivas pode preencher essa lacuna, facilitando o acesso a *insights* valiosos (Observatório de Educação, 2016; INEP, 2024). A incorporação de análises de tendências baseadas em regressão linear permite compreender o comportamento dos indicadores ao longo do tempo, apoiando o planejamento estratégico e a alocação eficiente de recursos (Revisa, 2025).

Além disso, a análise regional e temporal das taxas de abandono, aprovação e reprovação escolar, bem como da TDI, pode revelar disparidades significativas entre diferentes áreas geográficas do país (INEP, 2024; Jeduca, 2023). Identificar essas desigualdades é fundamental para promover políticas educacionais mais equitativas e direcionadas às necessidades específicas de cada região.

Este trabalho justifica-se, portanto, pela necessidade de transformar dados educacionais brutos em informações acessíveis e acionáveis, buscando contribuir para uma gestão educacional mais transparente, eficiente e orientada por evidências (Observatório de Educação, 2016; Observatório de Educação, 2021).

1.2 Objetivos

1.2.1 *Objetivo Geral*

O objetivo geral deste trabalho é realizar análises descritivas de indicadores educacionais brasileiros ao longo de séries históricas extensas, identificando tendências temporais e padrões regionais por meio de técnicas estatísticas. Como produto dessas análises, foi desenvolvido um *dashboard* interativo que consolida e disponibiliza os resultados de forma visual e intuitiva, permitindo que gestores, pesquisadores e formuladores de políticas públicas explorem os dados e informações identificados.

1.2.2 *Objetivos Específicos*

Os objetivos específicos deste trabalho são:

1. **Coletar e consolidar dados educacionais históricos:** Reunir e organizar dados de indicadores educacionais disponibilizados pelo INEP, incluindo métricas de abandono, aprovação e reprovação escolar (2007–2024) e Taxa de Distorção Idade-Série – TDI (2006–2024), agregando-os por código de município, estado e região.

2. **Analisar o comportamento temporal dos indicadores educacionais:** Investigar como as métricas de fluxo escolar e a TDI evoluíram ao longo das séries históricas, identificando tendências, padrões e variações regionais que possam informar políticas educacionais.
3. **Identificar tendências através de análises de regressão linear:** Implementar modelos de regressão linear utilizando os quatro indicadores educacionais selecionados (TDI, abandono, aprovação e reprovação) para identificar tendências nos dados históricos, avaliando a qualidade do ajuste por meio de métricas estatísticas como o coeficiente de determinação (R^2).
4. **Criar um *dashboard* interativo para visualização:** Desenvolver uma interface web utilizando React e Plotly que permita aos usuários explorar os dados de forma dinâmica, aplicar filtros por região, estado e período temporal, e visualizar tendências históricas por meio de gráficos interativos.
5. **Avaliar as tendências identificadas e disponibilizar resultados para gestão educacional:** Analisar os resultados das regressões lineares por meio de métricas estatísticas e interpretar os resultados à luz do contexto educacional brasileiro, disponibilizando as análises de forma acessível para subsidiar a formulação de políticas públicas e o direcionamento estratégico de recursos.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos técnicos essenciais para a compreensão deste trabalho. A estrutura é organizada da seguinte forma: A Seção 2.1 introduz os indicadores educacionais e sua relevância na análise de dados. A Seção 2.2 aborda o conceito de mineração de dados, com foco em sua importância para a ciência de dados. A Seção 2.3 trata do pipeline de dados, destacando sua função no processamento eficiente de grandes volumes de informações. A Seção 2.4 explora a visualização de dados e *dashboards* interativos, e suas aplicações no contexto educacional. A Seção 2.5 apresenta a análise estatística e regressão linear. Por fim, a Seção 2.6 descreve as tecnologias utilizadas no desenvolvimento do sistema.

2.1 Indicadores Educacionais

Indicadores educacionais são métricas cruciais para avaliar e monitorar o desempenho, a qualidade e a equidade dos sistemas de ensino. Eles abrangem diversos aspectos, como acesso e participação (taxa de matrícula e retenção), desempenho acadêmico (notas em avaliações padronizadas e taxa de conclusão), qualidade do ensino (relação aluno-professor e formação dos professores) e equidade e inclusão (desigualdade de desempenho e acesso à educação inclusiva).

Esses indicadores são fundamentais para o diagnóstico situacional, o monitoramento e a avaliação de políticas educacionais, a transparência e a prestação de contas, e o planejamento e a alocação de recursos. No entanto, sua efetividade pode ser afetada pela qualidade dos dados e pelas variações contextuais e culturais. A integração e análise desses indicadores, especialmente com o uso de técnicas de ciência de dados e análise descritiva, são essenciais para a formulação de estratégias que visem a melhoria contínua do sistema educacional e a redução das desigualdades.

2.2 Mineração de Dados

Mineração de dados é o processo de descoberta de padrões, correlações e anomalias significativas em grandes conjuntos de dados, utilizando técnicas de estatística, aprendizado de máquina e inteligência artificial. Essa prática é essencial para transformar dados brutos em informações úteis, permitindo que organizações tomem decisões informadas baseadas em evidências.

A mineração de dados emerge como um processo sistemático, interativo e iterativo de preparação e extração de conhecimentos a partir de grandes bases de dados. De acordo com

Castro e Ferrari, pode-se conceituar como:

O termo mineração de dados (MD) foi cunhado como alusão ao processo de mineração, uma vez que se explora uma base de dados (mina) usando o algoritmo (ferramentas) adequados para obter conhecimento (minerais preciosos). Os dados são símbolos ou signos não estruturados, sem significados, como valores em uma tabela, e a informação está contida nas descrições, agregando significado e utilidade aos dados, como o valor da temperatura do ar. Por fim, o conhecimento é algo que permite uma tomada de decisão para a agregação de valor. (Castro; Ferrari, 2017).

No contexto educacional, a mineração de dados pode ser utilizada para identificar fatores que influenciam o desempenho dos alunos, identificar padrões de comportamento e personalizar estratégias de ensino para diferentes perfis de estudantes. As etapas principais de um processo de mineração de dados incluem: seleção, pré-processamento, transformação, modelagem, avaliação e implementação.

2.3 Conceito de Pipeline de Dados

Embora atualmente exista um grande volume de dados sendo gerados, esses dados, em estado bruto, raramente têm alguma utilidade direta. Mecanismos são criados justamente para consumir e transformar esses dados passando por uma sucessão de operações, também conhecidas como pipeline de dados (Quemy, 2019). O pipeline de dados é uma sequência de processos que facilita a extração, transformação e carga (ETL) de dados de diversas fontes para um destino específico, onde podem ser armazenados, processados e analisados.

Os mesmos funcionam como uma cadeia com uma sequência de atividades simples e complexas, que realizam a manipulação desde a obtenção dos mesmos em diversas origens, até as transformações desses dados diante da finalidade desejada para o destino que envolve o armazenamento e aplicações em ferramentas de visualização e modelos de aprendizado de máquina (Raj *et al.*, 2020).

2.4 Visualização de Dados e Dashboards Interativos

2.4.1 Visualização de Dados

O uso de sistemas educacionais gera um crescimento exponencial da quantidade de dados a serem analisados, fornecendo uma oportunidade para novas formas de representação

destes dados. A visualização de dados trabalha na representação das informações em forma de gráficos, tabelas, imagens ou outras representações visuais para tornar a informação mais fácil de compreender e interpretar, sendo uma parte importante da análise de dados.

Segundo Grinstein e Ward (2002), os benefícios da utilização da visualização de dados incluem a capacidade de destacar tendências, relações e padrões em grandes quantidades de dados, ajudando os usuários a identificar informações valiosas e tomar decisões informadas. Existem vários tipos de visualizações de dados, incluindo gráficos de barras, gráficos de linha, gráficos de dispersão, mapas e gráficos de setores. Cada tipo de visualização é adequado para diferentes tipos de dados e objetivos, sendo importante escolher a visualização adequada para representar a informação de maneira clara e precisa.

A visualização de dados também pode ser combinada com tecnologias avançadas, como inteligência artificial e aprendizado de máquina, para criar soluções ainda mais poderosas e eficazes. Sistemas de visualização de dados podem ser integrados a bancos de dados e fontes de dados em tempo real para fornecer informações atualizadas (Unwin, 2020).

2.4.2 Dashboards Educacionais

Dashboards são ferramentas operacionais utilizadas para exibição visual de informações, com foco na análise dos principais dados, de modo a fornecer uma visão geral dos objetivos e o monitoramento das atividades (Few, 2006; Friendly; Wainer, 2021). Com a velocidade, variabilidade e grande volume de dados provenientes das inúmeras atividades praticadas em um sistema, naturalmente uma abundância de dados é gerada.

Os *dashboards* podem exibir dados de desempenho concretos com visualizações em uma única plataforma, permitindo que usuários acessem e sintetizem informações de forma eficiente, visando melhorar o desempenho, auxiliando a interpretação dos dados e integrando dados de várias fontes de avaliação (Boscardin *et al.*, 2018; Ellaway *et al.*, 2014).

Os *dashboards* educacionais desempenham um papel crucial na simplificação da análise de dados complexos no contexto educacional. Entre os diversos tipos de gráficos empregados para representar informações, destacam-se:

- **Gráficos de barras:** utilizados para visualizar comparativamente valores de diferentes categorias, como por exemplo médias regionais, valores máximos e mínimos de indicadores educacionais, oferecendo uma representação visual clara das discrepâncias entre conjuntos de dados discretos.

- **Gráficos de linha:** particularmente eficazes na representação de tendências ao longo do tempo, proporcionando informações dinâmicas sobre padrões e variações temporais.
- **Mapas interativos:** representações geográficas que permitem visualizar dados regionalizados, possibilitando análises comparativas entre estados, regiões e municípios de forma espacial e intuitiva. No contexto educacional brasileiro, mapas interativos são particularmente úteis para identificar disparidades regionais em indicadores educacionais, facilitando a formulação de políticas públicas direcionadas a áreas específicas.

A integração desses diversos tipos de gráficos em *dashboards* educacionais proporciona uma abordagem abrangente na apresentação visual de dados educacionais, facilitando a análise detalhada no contexto da gestão educacional.

2.5 Análise Estatística e Regressão Linear

2.5.1 Estatística Descritiva

A estatística descritiva consiste em métodos para organizar, resumir e apresentar dados de forma informativa, permitindo a compreensão das características principais de um conjunto de dados (Bussab; Morettin, 2017). No contexto educacional, técnicas estatísticas descritivas incluem medidas de tendência central (média, mediana, moda), medida de dispersão (desvio padrão, variância) e análises de distribuição de frequências.

Essas métricas são fundamentais para caracterizar o comportamento de indicadores educacionais ao longo de séries históricas, identificando valores típicos, variabilidade e padrões gerais nos dados. A análise estatística descritiva constitui a base para interpretações mais profundas sobre a evolução temporal de indicadores como taxas de aprovação, reprovação, abandono e distorção idade-série.

A análise descritiva representa o foco deste trabalho, permitindo compreender o comportamento histórico dos indicadores educacionais por meio de visualizações, estatísticas e análises comparativas entre regiões, estados e períodos temporais. Diferentemente de abordagens preditivas, que buscam projetar valores futuros, a análise descritiva concentra-se em extrair *insights* sobre o que já ocorreu, fornecendo uma base para compreensão dos dados e apoiando a tomada de decisões baseadas em evidências históricas (Hair *et al.*, 2009).

2.5.2 Regressão Linear

A regressão linear é uma técnica estatística utilizada para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes, permitindo a identificação de tendências e a análise de como as variáveis se relacionam ao longo do tempo (Montgomery *et al.*, 2012). No contexto de análise de séries temporais educacionais, a regressão linear pode ser aplicada para identificar tendências de crescimento ou declínio em indicadores ao longo dos anos.

O modelo de regressão linear simples pode ser expresso pela equação:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (2.1)$$

onde y representa a variável dependente (por exemplo, Taxa de Distorção Idade-Série ou taxa de aprovação), x representa a variável independente (por exemplo, tempo em anos), β_0 é o intercepto, β_1 é o coeficiente angular (taxa de variação) e ε representa o erro aleatório.

Neste trabalho, a regressão linear é utilizada como ferramenta de análise descritiva para identificar tendências históricas, não sendo empregada para realizar previsões de valores futuros. A reta de regressão serve como representação visual da direção geral dos dados ao longo do tempo, permitindo identificar se determinado indicador apresentou crescimento, declínio ou estabilidade no período analisado (ver Equação 2.1).

2.5.3 Coeficiente de Determinação (R^2)

O coeficiente de determinação, representado por R^2 , é uma métrica estatística que indica a proporção da variância na variável dependente que é explicada pelas variáveis independentes no modelo de regressão (Montgomery *et al.*, 2012). Matematicamente, o R^2 é definido como:

$$R^2 = 1 - \frac{SQR}{SQT} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.2)$$

onde:

- y_i representa o valor observado do indicador no ano i
- \hat{y}_i representa o valor predito pelo modelo de regressão linear para o ano i
- \bar{y} representa a média dos valores observados
- SQR (Soma dos Quadrados dos Resíduos) mede a variação não explicada pelo modelo

– *SQT* (Soma dos Quadrados Totais) mede a variação total dos dados em relação à média

O valor de R^2 varia entre 0 e 1, onde valores próximos a 1 indicam que o modelo explica uma grande proporção da variabilidade dos dados, enquanto valores próximos a 0 indicam baixo poder explicativo (UFPR, 2024). Um R^2 de 0,85, por exemplo, significa que 85% da variância no indicador é explicada pela variável tempo, e apenas 15% se deve a flutuações não capturadas pelo modelo linear.

No contexto de análise de tendências em indicadores educacionais, o R^2 permite avaliar a qualidade do ajuste linear aos dados históricos, indicando se a tendência identificada é consistente ao longo do tempo (Morettin; Toloi, 2006; Reis; Menezes, 2017). Valores elevados de R^2 sugerem que a tendência linear captura adequadamente o comportamento do indicador, enquanto valores baixos podem indicar a presença de variações não lineares ou fatores externos que influenciam o indicador de forma irregular (Asimov Academy, 2025; Costa *et al.*, 2017).

É importante ressaltar que, embora o R^2 forneça uma medida quantitativa da qualidade do ajuste da reta aos dados históricos, sua utilização neste trabalho limita-se à caracterização descritiva das tendências observadas no passado. O R^2 não é empregado aqui com propósito preditivo, mas sim como indicador da consistência e força da tendência histórica identificada, auxiliando na interpretação de padrões temporais já estabelecidos nos dados educacionais (Cortina, 2025; Leme Tec, 2024).

2.6 Tecnologias para Desenvolvimento do Sistema

2.6.1 *React*

React é uma biblioteca JavaScript de código aberto, criada e mantida pela Meta (anteriormente Facebook), lançada em 2013 (DIO, 2025). A biblioteca é amplamente utilizada para construir interfaces de usuário (*user interfaces* – UI) baseadas em componentes reutilizáveis e eficientes. React utiliza o conceito de *Virtual DOM* (*Document Object Model* virtual) para otimizar a renderização e melhorar o desempenho de aplicações web, sendo uma das tecnologias mais adotadas no desenvolvimento front-end atual (DIO, 2025).

A principal característica do React é sua arquitetura baseada em componentes, onde cada componente é uma unidade independente e reutilizável que encapsula sua própria lógica e apresentação. Essa abordagem facilita a manutenção do código, promove a reutilização e permite o desenvolvimento de aplicações complexas de forma mais organizada.

2.6.2 *Plotly*

Plotly é uma biblioteca JavaScript baseada em D3.js (*Data-Driven Documents*) que permite a criação de gráficos interativos com funcionalidades avançadas (Asimov Academy, 2025). A biblioteca oferece recursos como zoom, filtragem, *hover* (exibição de informações ao passar o mouse) e exportação para formatos web, sendo amplamente utilizada na construção de *dashboards* e aplicações que requerem exploração dinâmica de dados.

O Plotly.js é conhecido por suas capacidades de interatividade e oferece suporte a mais de 30 tipos de gráficos, permitindo que usuários interajam diretamente com as visualizações (Asimov Academy, 2025; DIO, 2025). Essa interatividade torna a análise de dados mais intuitiva e detalhada, possibilitando que gestores educacionais explorem tendências, identifiquem padrões e realizem análises comparativas de forma eficiente.

A combinação de Python no back-end com React e Plotly no front-end permite o desenvolvimento de *dashboards* educacionais modernos e responsivos, onde a camada de processamento de dados (Python) se integra com a interface interativa (React + Plotly), proporcionando uma experiência de usuário eficaz.

2.6.3 *Python*

Python é uma linguagem de programação de alto nível, interpretada e de propósito geral, amplamente utilizada em ciência de dados, análise estatística e desenvolvimento web (Rossum; Drake, 2009). Criada por Guido van Rossum e lançada em 1991, Python destaca-se por sua sintaxe clara e legível, facilitando o desenvolvimento e a manutenção de código (Python Software Foundation, 2024).

No desenvolvimento de *dashboards* educacionais, Python é frequentemente utilizado no back-end para processamento de dados, cálculo de métricas estatísticas, execução de modelos de regressão e preparação de dados para consumo via API (*Application Programming Interface*) pelo front-end

3 TRABALHOS RELACIONADOS

Neste capítulo, são apresentados trabalhos que guardam similaridades com o projeto aqui estabelecido, destacando-se os aspectos metodológicos, técnicas empregadas e domínios de aplicação. A análise comparativa permite identificar tanto as contribuições dos trabalhos anteriores quanto os aspectos em que este projeto se diferencia, seja pela abordagem, pelo foco em análise descritiva ou pela aplicação no contexto educacional brasileiro.

3.1 *From E-budgeting to Smart Budgeting: Exploring the Potential of Artificial Intelligence in Government Decision-Making for Resource Allocation*

O estudo de Valle-Cruz *et al.* (2022) aborda a aplicação da inteligência artificial (IA) na alocação de recursos públicos, uma área ainda pouco explorada na literatura. A pesquisa investiga em que medida técnicas de IA podem contribuir para uma distribuição eficiente dos gastos públicos, visando maximizar o Produto Interno Bruto (PIB), reduzir a inflação e diminuir o índice de Gini. Utilizando dados abertos do Banco Mundial de 1960 a 2019, o estudo emprega redes neurais do tipo perceptrons multicamadas e algoritmos genéticos multiobjetivo para analisar como entradas de orçamento são processadas para gerar resultados econômicos, políticos e sociais.

Os autores revisam a literatura recente sobre o uso de IA na tomada de decisões governamentais, com especial atenção para a análise de como os gastos públicos impactam os resultados sociais e econômicos. A abordagem metodológica segue o paradigma de modelos algorítmicos, que trata o mecanismo dos dados como desconhecido, ao invés de utilizar modelos estocásticos tradicionais.

Os resultados do estudo indicam que a aplicação de técnicas de IA pode melhorar a precisão na identificação de relações não lineares entre despesas orçamentárias e resultados econômicos, possibilitando uma melhor compreensão da importância de diferentes categorias de gastos. O modelo proposto demonstrou ser eficaz em identificar padrões que podem auxiliar na otimização da alocação de recursos públicos.

Relação com o presente trabalho:

Embora o estudo de Valle-Cruz *et al.* (2022) utilize técnicas preditivas de IA, enquanto este trabalho adota uma abordagem descritiva, ambos compartilham o objetivo de utilizar dados públicos para informar decisões governamentais baseadas em evidências. A utilização de

análises quantitativas para compreender padrões em indicadores sociais (econômicos no caso de Valle-Cruz *et al.* (2022), educacionais no presente trabalho) representa um ponto de convergência metodológica. A diferença fundamental reside no fato de que este trabalho não busca prever resultados futuros ou otimizar alocações, mas sim **caracterizar tendências históricas** e disponibilizá-las por meio de um *dashboard* interativo para apoiar gestores educacionais.

3.2 *A Decision Model for Municipal Resources Management*

O artigo de Thesari *et al.* (2019) propõe um modelo para apoiar gestores locais na otimização do orçamento público municipal, utilizando uma integração de métodos. O modelo foi desenvolvido para suprir a lacuna relacionada à definição de pesos em avaliações subjetivas, geralmente problemáticas, auxiliando na tomada de decisão quanto à distribuição do orçamento e identificando a importância dos setores nos governos locais baseada em dados históricos.

A metodologia seguiu três etapas principais: exploração das características dos setores locais representados pelos departamentos municipais e coleta de dados usando séries temporais; cálculo dos pesos referentes à importância de cada departamento municipal pelo método UTASTAR com base nos dados históricos; e formulação de uma função objetivo usando programação linear e restrições baseadas em especificações legais, resultando em uma projeção otimizada para a distribuição do orçamento público.

Os resultados indicaram que o modelo é eficiente na definição de pesos considerando o comportamento das preferências por meio de dados históricos e auxiliando na otimização dos recursos públicos locais, além de garantir conformidade com a legislação. Ele também é capaz de prever ou projetar valores futuros disponíveis no orçamento.

Relação com o presente trabalho:

A utilização de séries temporais e dados históricos para análise de indicadores públicos é um ponto de convergência entre o trabalho de Thesari *et al.* (2019) e este projeto. Ambos reconhecem a importância de compreender o comportamento histórico de indicadores para apoiar decisões governamentais. No entanto, enquanto Thesari *et al.* (2019) focam em **otimização e alocação prescritiva** de recursos municipais por meio de programação linear, este trabalho concentra-se em **análise descritiva de tendências** educacionais e na disponibilização dessas análises por meio de visualizações interativas. A abordagem deste trabalho não prescreve alocações ótimas, mas fornece subsídios informacionais para que gestores compreendam a evolução histórica dos indicadores e identifiquem áreas que requerem atenção.

3.3 *An Application of Machine Learning to Study Utilities Expenses in the Brazilian Navy*

O trabalho apresentado investiga a aplicação de técnicas de aprendizado de máquina para apoiar o processo de tomada de decisão em relação às despesas das instalações da Marinha Brasileira. O estudo foi conduzido em um contexto de variação significativa de custos, envolvendo mais de 350 instalações com atividades diversas, desde operações militares até funções administrativas e de suporte. A abordagem proposta utilizou modelos preditivos para analisar atributos dessas instalações, enfrentando desafios como limitações de dados e a necessidade de compreender a natureza diferenciada das despesas.

Os modelos foram treinados para classificar e prever os custos, e os resultados obtidos foram aplicados para otimizar a gestão dos recursos da organização, proporcionando informações valiosas para melhorar a eficiência e a alocação de recursos. O estudo ressalta a importância de uma abordagem baseada em dados para a gestão e otimização de recursos em diferentes contextos organizacionais.

Relação com o presente trabalho:

Embora o estudo sobre as despesas da Marinha Brasileira empregue técnicas de aprendizado de máquina com foco preditivo, ele compartilha com este trabalho o desafio de lidar com dados complexos, heterogêneos e com limitações de completude. A necessidade de realizar pré-processamento extensivo, limpeza de dados e engenharia de atributos são aspectos metodológicos comuns a ambos os trabalhos. A principal diferença reside no objetivo final: enquanto o estudo da Marinha busca **prever e classificar** despesas futuras, este trabalho busca **descrever e caracterizar** tendências históricas em indicadores educacionais. Ambos os trabalhos demonstram como análises baseadas em dados podem informar gestão de recursos públicos, mas em contextos e com propósitos distintos.

3.4 *Dashboard Interativo para Análise de Indicadores Educacionais: Uma Abordagem Baseada em Dados Abertos*

A dissertação de Menezes (2023) apresenta o desenvolvimento de um *dashboard* interativo para análise de indicadores educacionais brasileiros, utilizando dados abertos disponibilizados pelo INEP. O trabalho teve como objetivo criar uma ferramenta que facilitasse a visualização e interpretação de indicadores como IDEB, taxas de rendimento escolar e adequação da formação docente, permitindo análises comparativas entre municípios, estados e regiões.

A metodologia empregada envolveu a coleta automatizada de dados do INEP, pré-processamento utilizando Python e Pandas, desenvolvimento de visualizações com bibliotecas como Plotly e Matplotlib, e implementação de uma aplicação *web* interativa. O *dashboard* desenvolvido oferece funcionalidades de filtragem por região, estado, município e período temporal, permitindo que gestores e pesquisadores explorem os dados de forma intuitiva.

Os resultados demonstraram a viabilidade de consolidar múltiplas fontes de dados educacionais em uma interface unificada, facilitando análises que antes exigiam manipulação manual de diversos arquivos e planilhas. O trabalho evidenciou a importância de ferramentas de visualização para transformar dados brutos em *insights* acionáveis no contexto da gestão educacional.

Relação com o presente trabalho:

Este trabalho apresenta forte alinhamento metodológico e conceitual com a dissertação de Menezes (2023), compartilhando objetivos similares de desenvolver ferramentas interativas para análise de indicadores educacionais brasileiros baseadas em dados abertos. Ambos os trabalhos enfatizam a importância da visualização de dados e da disponibilização de interfaces intuitivas para apoiar gestores educacionais.

As principais convergências incluem, especialmente, o desenvolvimento de *dashboards* interativos com funcionalidades de filtragem e emprego de Python para processamento de dados e bibliotecas de visualização para criação de gráficos, focando em facilitar a interpretação de indicadores educacionais.

No entanto, as diferenças principais residem em quatro aspectos. Primeiro, este trabalho incorpora análises de regressão linear e cálculo do coeficiente de determinação (R^2) para caracterizar tendências históricas de forma quantitativa, enquanto Menezes (2023) foca em visualizações descritivas sem modelagem estatística formal. Segundo, este trabalho utiliza mapas choropleth interativos para visualização espacial dos indicadores por estado, recurso não presente em Menezes (2023), permitindo identificar padrões geográficos e disparidades regionais. Terceiro, a arquitetura tecnológica deste trabalho utiliza React com TypeScript no frontend e FastAPI no backend, diferenciando-se das escolhas tecnológicas de Menezes (2023), que empregou PHP e Python.

Por fim, este trabalho realiza uma análise de completude e consistência dos dados disponibilizados pelo INEP. Durante a fase exploratória, foram identificadas lacunas em diversos indicadores, como dados ausentes ou incompletos em determinados anos e municípios. Essa

análise crítica motivou a seleção de quatro indicadores — TDI, Aprovação, Reprovação e Abandono — que apresentavam maior consistência temporal e cobertura geográfica, garantindo maior confiabilidade nas análises apresentadas.

Dessa forma, este trabalho pode ser considerado complementar à dissertação de Menezes (2023), aprofundando aspectos de análise estatística de tendências e oferecendo uma perspectiva adicional sobre a evolução temporal dos indicadores educacionais brasileiros.

3.5 Síntese Comparativa

O Quadro 1 apresenta uma síntese comparativa dos trabalhos relacionados, destacando as principais características metodológicas, objetivos e relações com o presente trabalho.

Quadro 1 – Síntese comparativa dos trabalhos relacionados

Trabalho	Domínio	Abordagem	Técnicas	Objetivo Principal	Relação com este TCC
Vallecruz et al. (2022)	Orçamento Público	Preditiva	Redes Neurais, Algoritmos Genéticos	Otimizar alocação de recursos	Uso de dados públicos para decisões baseadas em evidências
Thesari et al. (2019)	Gestão Municipal	Prescritiva	UTASTAR, Programação Linear	Otimizar distribuição orçamentária	Análise de séries temporais e dados históricos
Marinha Brasileira	Gestão de Despesas	Preditiva	Aprendizado de Máquina	Prever e classificar custos	Desafios de dados complexos e heterogêneos
Menezes (2023)	Educação	Descritiva	Visualização de Dados	<i>Dashboard</i> interativo educacional	Forte alinhamento metodológico e conceitual
Este Trabalho	Educação	Descritiva	Regressão Linear	Caracterizar tendências históricas via <i>dashboard</i>	—

Fonte: Elaborado pela autora.

Os trabalhos relacionados apresentados demonstram a relevância de abordagens baseadas em dados para análise de indicadores públicos em diversos domínios. Embora vários dos trabalhos empreguem técnicas preditivas e prescritivas, este trabalho diferencia-se ao focar em **análise descritiva robusta** de tendências históricas, complementada por um *dashboard*

interativo que democratiza o acesso aos *insights* derivados das análises.

4 METODOLOGIA

Este capítulo descreve de forma detalhada os procedimentos metodológicos adotados para o desenvolvimento deste trabalho, abrangendo desde a coleta e preparação dos dados educacionais até a implementação do *dashboard* interativo. A metodologia seguiu uma abordagem sistemática e iterativa, dividida em etapas sequenciais, cada uma com objetivos específicos que contribuíram para o alcance dos resultados finais. Este trabalho caracteriza-se por sua natureza aplicada e exploratória, com foco em análise descritiva de dados históricos, utilizando técnicas de estatística, visualização de dados e desenvolvimento web para criar uma ferramenta de suporte à gestão educacional baseada em evidências.

4.1 Natureza da Pesquisa

A natureza deste projeto se define por ser uma pesquisa aplicada, com a finalidade de gerar conhecimentos para aplicação prática no contexto educacional brasileiro. A pesquisa aplicada busca resolver um problema específico, caracterizando-se por ser orientada diretamente para a busca de soluções (Brown, 2015). Neste caso, o problema identificado foi a dificuldade de gestores educacionais, pesquisadores e formuladores de políticas públicas em analisar e interpretar grandes volumes de dados históricos de indicadores educacionais de forma intuitiva, acessível e visual.

A pesquisa aplicada justifica-se pela necessidade de transformar dados brutos disponibilizados por órgãos governamentais, especificamente pelo INEP, em informações acionáveis que possam orientar decisões estratégicas na área educacional. A natureza aplicada do trabalho manifesta-se na produção de um artefato tecnológico concreto – o *dashboard* interativo – que pode ser utilizado por diferentes *stakeholders* do sistema educacional para explorar indicadores, identificar tendências históricas e compreender disparidades regionais.

4.2 Abordagem da Pesquisa

Para alcançar os objetivos desta pesquisa, optou-se por adotar uma abordagem descritiva e exploratória. A abordagem descritiva visa apresentar características específicas de uma população ou fenômeno, permitindo estabelecer correlações entre variáveis e entender melhor sua natureza ao longo do tempo. No contexto deste trabalho, a abordagem descritiva é aplicada para caracterizar o comportamento de indicadores educacionais brasileiros ao longo

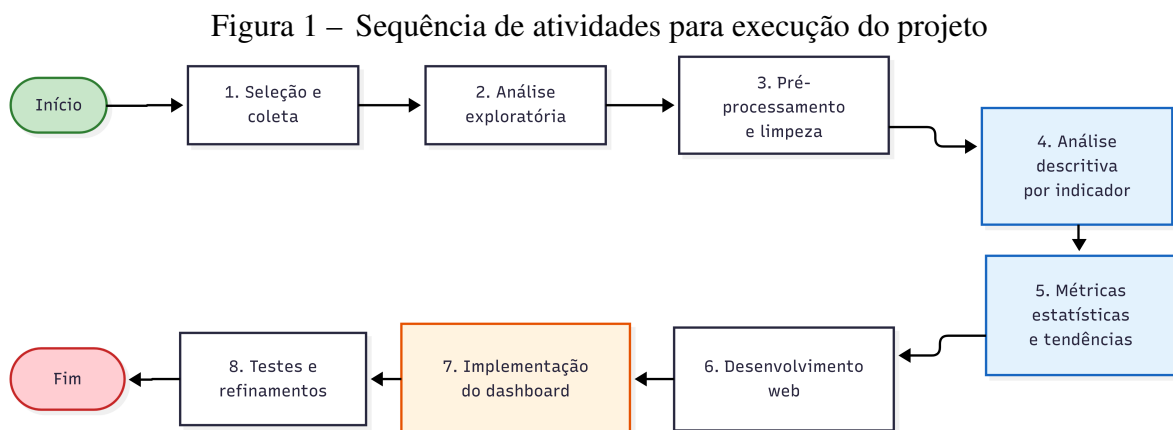
de séries históricas extensas, identificando padrões temporais, variações regionais e tendências gerais nos dados.

A análise exploratória de dados (EDA – *Exploratory Data Analysis*) foi essencial para identificar padrões, tendências e características nos indicadores educacionais históricos, fornecendo subsídios para o desenvolvimento das visualizações implementadas no *dashboard* (Tukey, 1977; Hair *et al.*, 2009). A exploração iterativa dos dados permitiu a descoberta de *insights* sobre a evolução dos indicadores ao longo do tempo, revelando melhorias significativas em alguns indicadores e desigualdades regionais persistentes.

É importante ressaltar que este trabalho não adota uma abordagem preditiva, embora utilize técnicas de regressão linear e coeficiente de determinação (R^2). A regressão linear é empregada exclusivamente como ferramenta de análise descritiva para identificar e quantificar tendências históricas, sem o objetivo de realizar previsões de valores futuros. O foco central do trabalho está na análise descritiva, que busca compreender o que já aconteceu, e não em projetar cenários futuros.

4.3 Etapas Metodológicas

O desenvolvimento deste trabalho seguiu uma sequência de etapas metodológicas estruturadas e interdependentes, cada uma contribuindo para a construção progressiva do conhecimento sobre os indicadores educacionais e para o desenvolvimento do *dashboard* interativo. A Figura 1 apresenta uma visão geral do fluxo metodológico adotado, ilustrando a sequência de atividades desde a coleta de dados até a implementação final do sistema. O diagrama presente na Figura 1 ilustra o fluxo do trabalho.



Fonte: Elaborado pela autora.

As etapas metodológicas foram organizadas da seguinte forma: (1) Seleção e coleta do conjunto de dados; (2) Análise exploratória inicial e avaliação de completude; (3) Pré-processamento, limpeza e transformação dos dados; (4) Análise descritiva aprofundada por indicador; (5) Cálculo de métricas estatísticas e análise de tendências; (6) Desenvolvimento de visualizações; (7) Implementação do *dashboard* interativo; e (8) Testes e refinamentos do sistema.

4.3.1 Seleção e Coleta do Conjunto de Dados

A primeira etapa metodológica consistiu na identificação, seleção e coleta dos dados educacionais disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), vinculado ao Ministério da Educação (MEC). O INEP é a principal entidade responsável por coletar, sistematizar e disponibilizar dados sobre a educação brasileira, oferecendo acesso público a uma vasta gama de indicadores educacionais por meio de iniciativas como o Inep Data e o Painel Educacional.

4.3.1.1 Fonte e Acesso aos Dados

Os dados foram obtidos por meio da plataforma oficial do INEP¹, que disponibiliza microdados e indicadores educacionais em formato aberto, em conformidade com a Lei de Acesso à Informação (Lei nº 12.527/2011) (BRASIL, 2011). A disponibilização pública desses dados visa promover transparência, permitir o controle social e fomentar pesquisas acadêmicas e aplicadas na área educacional.

4.3.1.2 Processo de Coleta Automatizada

Para garantir eficiência e reprodutibilidade no processo de coleta, foram desenvolvidos scripts automatizados em Python utilizando bibliotecas especializadas. O processo de automatização envolveu as seguintes etapas:

- **Download automatizado de arquivos:** Utilizando as bibliotecas `requests` e `BeautifulSoup`, foram desenvolvidos scripts que acessam o portal do INEP, identificam os arquivos ZIP contendo os indicadores educacionais de interesse e realizam o download automatizado.
- **Extração de arquivos compactados:** Os arquivos baixados, geralmente em formato ZIP,

¹ Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos>

foram extraídos utilizando a biblioteca `zipfile` do Python, que descompacta os arquivos e organiza seu conteúdo em diretórios estruturados por ano.

- **Organização hierárquica:** Os dados foram organizados em uma estrutura de pastas hierárquica, onde cada ano possui seu próprio diretório contendo os arquivos CSV correspondentes aos diferentes indicadores educacionais disponíveis para aquele período.

4.3.1.3 Indicadores Educacionais Considerados Inicialmente

Na fase inicial da pesquisa, foram considerados 14 indicadores educacionais diferentes, cujas descrições detalhadas estão apresentadas no Quadro 2. Esses indicadores abrangem diversas dimensões do sistema educacional brasileiro, incluindo:

Quadro 2 – Descrição dos indicadores utilizados.

Indicador	Descrição
Adequação da Formação Docente (AFD)	Avalia se a formação inicial dos professores está adequada às disciplinas que eles lecionam.
Complexidade de Gestão da Escola (ICG)	Avalia o nível de dificuldade de administração.
Esforço Docente (IED)	Avalia a carga de trabalho dos professores na educação.
Índice de Desenvolvimento da Educação Básica (IDEB)	Mede a qualidade do ensino nas escolas brasileiras.
Média de Alunos por Turma (ATU)	Mede a quantidade média de alunos em cada turma, sendo um indicador da densidade e do ambiente de sala de aula.
Média de Horas-aula Diária (HAD)	Avalia o tempo médio de instrução diária, refletindo a intensidade do ensino oferecido aos alunos.
Indicador de Nível Socioeconômico (INSe)	Verifica o impacto do contexto socioeconômico nos resultados educacionais.
Percentual de Docentes com Curso Superior	Mede a proporção de professores que possuem formação superior, um fator importante para a qualidade do ensino.
Regularidade do Corpo Docente	Avalia a estabilidade do corpo docente nas escolas de educação básica ao longo de um período de cinco anos.
Taxas de Distorção Idade-série	Mede o percentual de estudantes matriculados em determinada série que estão com uma idade superior à idade ideal para essa série.
Taxas de Não-resposta	Mede o percentual de alunos para os quais não há informações registradas sobre seu rendimento escolar.
Taxas de Rendimento Escolar	Avalia o desempenho dos alunos nas escolas, incluindo aprovação, reprovação e abandono.

Fonte: Elaborado pela autora.

4.3.1.4 Seleção Final dos Indicadores

Após a coleta inicial dos 14 indicadores, foi realizada uma análise exploratória para avaliar a viabilidade de utilização de cada indicador em análises longitudinais. Esta análise revelou que diversos indicadores apresentavam disponibilidade irregular ao longo da série temporal, com anos específicos sem cobertura completa ou municípios sem registros, comprometendo a consistência necessária para análises de tendências e comparações entre períodos.

Com base nos critérios de completude e relevância para a análise da qualidade educacional, foram selecionados quatro indicadores principais para as análises aprofundadas:

- **Taxa de Distorção Idade-Série (TDI):** dados de 2006 a 2024 (19 anos de série histórica).
- **Taxa de Abandono:** dados de 2007 a 2024 (18 anos de série histórica).
- **Taxa de Aprovação:** dados de 2007 a 2024 (18 anos de série histórica).
- **Taxa de Reprovação:** dados de 2007 a 2024 (18 anos de série histórica).

A seleção desses quatro indicadores baseou-se em três critérios principais: (1) disponibilidade de séries históricas longas e consistentes; (2) baixo percentual de dados faltantes (inferior a 30%); e (3) relevância direta para a composição do Índice de Desenvolvimento da Educação Básica (IDEB), que é o principal indicador sintético da qualidade educacional brasileira (INEP, 2024).

4.3.1.5 Nível de Agregação dos Dados

Os dados foram coletados no nível de agregação municipal, utilizando o código de município do Instituto Brasileiro de Geografia e Estatística (IBGE) como chave primária de identificação. Esse nível de granularidade permite análises detalhadas sobre disparidades locais e facilita a comparação entre municípios, estados e regiões do país.

A partir dos códigos municipais, foram adicionadas informações complementares de estado (Unidade Federativa – UF) e região geográfica (Norte, Nordeste, Centro-Oeste, Sudeste e Sul), permitindo agregações e análises em diferentes níveis territoriais. Essa estrutura de dados hierárquica facilita a exploração dos indicadores tanto no nível micro (municipal) quanto nos níveis meso (estadual) e macro (regional).

4.3.1.6 *Total de Municípios Analisados*

As análises abrangeram um total de 5.562 municípios brasileiros, representando a quase totalidade dos municípios do país que possuem sistemas educacionais ativos e dados reportados ao INEP (IBGE, 2024). Essa cobertura abrangente garante a representatividade nacional das análises e permite identificar padrões gerais, bem como particularidades regionais e locais.

4.3.2 *Análise Exploratória Inicial e Avaliação de Completude*

Após a coleta inicial dos dados, foi realizada uma análise exploratória criteriosa para avaliar a qualidade, completude e adequação dos dados antes de prosseguir com análises mais aprofundadas (Hair *et al.*, 2009). Esta etapa foi fundamental para identificar limitações nos dados e orientar a seleção final dos indicadores a serem utilizados.

4.3.2.1 *Processo de Consolidação dos Indicadores*

O primeiro passo da análise exploratória foi a consolidação dos diferentes indicadores educacionais em um único conjunto de dados integrado. Utilizando a biblioteca Pandas do Python, foram realizadas operações de *merge* (junção) entre os diversos arquivos CSV, utilizando como chaves de junção o código de município e o ano (McKinney, 2010).

A operação de *merge* foi realizada utilizando a estratégia de *full outer join*, que preserva todos os registros de todas as tabelas, mesmo quando não há correspondência entre as chaves. Essa abordagem foi escolhida para evitar perda de informações e para permitir uma avaliação completa da disponibilidade de dados para cada município e ano.

4.3.2.2 *Análise de Dados Faltantes*

Com os dados consolidados, foi realizada uma análise sistemática de dados faltantes (*missing data*), que consiste na identificação e quantificação de células vazias ou com valores ausentes no conjunto de dados (Hair *et al.*, 2009). Para cada variável (indicador educacional), foram calculados:

- **Número absoluto de valores faltantes:** contagem do total de registros sem informação para aquele indicador.
- **Percentual de valores faltantes:** proporção de registros ausentes em relação ao total de

registros possíveis.

- **Padrão de ausência:** identificação se os dados faltantes estão concentrados em períodos específicos, regiões ou tipos de municípios.

A análise revelou que, dos 14 indicadores inicialmente considerados, apenas quatro apresentavam menos de 30% de dados faltantes, limite considerado aceitável para análises descritivas de séries temporais segundo Hair *et al.* (2009). Os demais indicadores apresentavam taxas de ausência superiores a 50%, inviabilizando análises longitudinais confiáveis.

4.3.2.3 *Crítérios de Seleção dos Indicadores*

Com base nos resultados da análise de completude, foram estabelecidos os seguintes critérios para seleção dos indicadores que seriam utilizados nas etapas subsequentes:

- **Completude mínima:** percentual de dados faltantes inferior a 30%.
- **Série histórica:** disponibilidade de dados para pelo menos 15 anos consecutivos ou semi-consecutivos.
- **Relevância teórica:** importância do indicador para avaliação da qualidade educacional, especialmente sua relação com o IDEB.
- **Cobertura geográfica:** disponibilidade de dados para a maioria dos municípios brasileiros.

Os quatro indicadores selecionados (TDI, Abandono, Aprovação e Reprovação) atenderam a todos esses critérios, apresentando séries históricas robustas e cobertura geográfica abrangente.

4.3.2.4 *Estatísticas Descritivas Iniciais*

Durante a análise exploratória inicial, foram calculadas estatísticas descritivas básicas para os quatro indicadores selecionados, incluindo:

- **Média geral:** valor médio do indicador considerando todos os municípios e anos.
- **Desvio padrão:** medida de dispersão dos valores em torno da média.
- **Valores mínimos e máximos:** identificação dos extremos da distribuição.

Essas estatísticas preliminares forneceram uma compreensão inicial do comportamento dos indicadores e orientaram as análises subsequentes.

4.3.2.5 *Validação de Intervalos*

Os valores dos indicadores foram validados para garantir que estavam dentro dos intervalos esperados. Especificamente:

- **Taxas percentuais:** verificação se os valores das taxas de Abandono, Aprovação e Reprovação estavam entre 0% e 100%.
- **TDI:** verificação se os valores da Taxa de Distorção Idade-Série estavam entre 0% e 100%.

A análise não identificou valores fora dos intervalos estabelecidos. Todos os registros apresentaram valores dentro dos limites plausíveis, não sendo necessária a remoção de dados durante esta etapa de validação.

4.3.2.6 *Adição de Variáveis Geográficas*

Para facilitar análises regionalizadas, foram adicionadas variáveis geográficas ao conjunto de dados:

- **Código UF:** código numérico da Unidade Federativa (estado).
- **Nome UF:** nome completo do estado.
- **Sigla UF:** sigla de duas letras do estado (ex.: CE, SP, BA).
- **Região:** região geográfica do Brasil (Norte, Nordeste, Centro-Oeste, Sudeste, Sul).
- **Código região:** código numérico da região (1 a 5).

Essas variáveis foram adicionadas por meio de operações de *join* com tabelas auxiliares contendo o mapeamento entre códigos de município e informações geográficas.

4.3.2.7 *Agregação Temporal*

Os dados foram organizados em uma estrutura longitudinal, onde cada linha representa um município em um ano específico. Essa estrutura, conhecida como formato *long* ou formato de painel, facilita análises de séries temporais e permite o cálculo de tendências ao longo do tempo (Morettin; Toloi, 2006).

4.3.2.8 *Exportação dos Dados Limpos*

Após todas as etapas de pré-processamento, os dados limpos e estruturados foram exportados para novos arquivos CSV organizados, servindo como entrada para as etapas subsequentes de análise. Essa separação entre dados brutos e dados processados garante rastreabilidade.

bilidade e permite que o processo de pré-processamento seja reproduzido ou ajustado conforme necessário.

4.3.3 *Análise Descritiva Aprofundada dos Indicadores*

Com os dados pré-processados e organizados, foi iniciada a etapa de análise descritiva aprofundada de cada um dos quatro indicadores selecionados. Esta etapa teve como objetivo compreender o comportamento histórico dos indicadores, identificar tendências temporais e revelar disparidades regionais (Hair *et al.*, 2009).

4.3.3.1 *Análise da Taxa de Distorção Idade-Série (TDI)*

A Taxa de Distorção Idade-Série é um indicador educacional que mede a proporção de alunos com dois ou mais anos de atraso escolar em relação à série que deveriam estar cursando, considerando sua idade. Valores elevados de TDI indicam problemas como repetência sucessiva, ingresso tardio no sistema educacional ou interrupções na trajetória escolar.

O objetivo da análise do TDI foi entender como essa métrica se comportou ao longo da série histórica de 2006 a 2024 e que tipo de informação essa evolução temporal poderia trazer para gestores educacionais e formuladores de políticas públicas.

4.3.3.1.1 Cálculo de Métricas Estatísticas

Para cada município e para cada indicador da análise, foram calculadas métricas estatísticas descritivas a partir da série histórica anual, considerando o primeiro e o último ano disponíveis. Essas métricas permitem quantificar o nível inicial, o nível final e a magnitude da mudança ao longo do período.

Especificamente, para cada indicador X , foram obtidas:

- **Valor inicial:** valor do indicador no primeiro ano da série, X_{inicial} .
- **Valor final:** valor do indicador no último ano da série, X_{final} .
- **Mudança absoluta:** diferença entre o valor final e o valor inicial, dada por

$$\Delta X = X_{\text{final}} - X_{\text{inicial}} \quad (4.1)$$

- **Mudança anual média:** variação média por ano, calculada como a mudança absoluta dividida pelo número de anos do período analisado.

- **Mudança acumulada média:** medida que sintetiza a tendência média de variação ao longo de toda a série histórica do indicador.
- **Variação percentual total:** percentual de mudança entre os valores inicial e final, calculado como

$$\text{Variação \%} = \frac{X_{\text{final}} - X_{\text{inicial}}}{X_{\text{inicial}}} \times 100\% \quad (4.2)$$

Essas métricas foram calculadas individualmente para todos os municípios e para todos os indicadores considerados, permitindo identificar casos de melhoria, estagnação ou piora ao longo do período analisado.

4.3.3.1.2 Análise Regional de Tendências

Para compreender se as mudanças nos indicadores foram homogêneas ou concentradas em determinadas regiões do país, foi realizada uma análise de tendências agrupadas por região geográfica. Para cada região (Norte, Nordeste, Centro-Oeste, Sudeste e Sul), foram calculadas as seguintes estatísticas agregadas:

- **Número de municípios:** quantidade de municípios analisados na região.
- **Mudança anual média:** variação média anual do indicador na região.
- **Variação percentual total:** percentual de mudança acumulada no período para a região.
- **Coefficiente de determinação médio (R^2):** média dos coeficientes de determinação dos municípios da região, indicando a consistência da tendência linear observada.
- **Desvio padrão:** medida de dispersão das mudanças entre municípios da mesma região, quantificando a heterogeneidade intrarregional.

Essa análise permitiu identificar as regiões com maiores mudanças no período e avaliar se existe convergência ou divergência entre regiões historicamente desiguais em termos de desempenho educacional.

4.3.3.2 Análise dos Indicadores de Fluxo Escolar (Abandono, Aprovação e Reprovação)

Após a análise aprofundada do TDI, o mesmo processo metodológico foi replicado para os três indicadores de fluxo escolar: Taxa de Abandono, Taxa de Aprovação e Taxa de Reprovação. Esses indicadores abrangem o período de 2007 a 2024 (18 anos de série histórica) e foram agregados com o código de município, adicionando-se as informações de estado e região.

O objetivo dessa análise foi entender como essas métricas se comportaram ao longo da série histórica e que tipo de informação essa evolução poderia trazer para políticas educacionais e gestão escolar.

4.3.4 Cálculo de Métricas Estatísticas e Análise de Tendências

Para quantificar e caracterizar as tendências identificadas visualmente nos dados, foram aplicadas técnicas de análise de tendências baseadas em regressão linear. É fundamental ressaltar que essa aplicação teve caráter exclusivamente descritivo, visando caracterizar matematicamente o comportamento histórico dos indicadores, e não realizar previsões de valores futuros.

4.3.4.1 Regressão Linear para Identificação de Tendências

Para cada município e para cada um dos quatro indicadores, foi ajustado um modelo de regressão linear simples, onde a variável independente é o tempo (ano) e a variável dependente é o valor do indicador naquele ano (Montgomery *et al.*, 2012). O modelo pode ser expresso pela equação:

$$y_t = \beta_0 + \beta_1 \cdot t + \varepsilon_t \quad (4.3)$$

onde y_t representa o valor do indicador no ano t , β_0 é o intercepto, β_1 é o coeficiente angular que indica a taxa de mudança anual, e ε_t representa o erro aleatório.

4.3.4.2 Coeficiente de Determinação (R^2)

Para cada regressão linear ajustada, foi calculado o coeficiente de determinação (R^2), que mede a proporção da variância no indicador que é explicada pela variável tempo (Montgomery *et al.*, 2012). O R^2 varia entre 0 e 1, onde:

- R^2 **próximo a 1** indica que a tendência linear explica bem o comportamento do indicador ao longo do tempo, sugerindo uma mudança consistente e sistemática.
- R^2 **próximo a 0** indica que há muita variabilidade não explicada pelo modelo linear, sugerindo que o indicador flutua de forma irregular ou não segue uma tendência linear clara.

4.3.4.2.1 Interpretação Descritiva do R^2

Neste trabalho, o R^2 é interpretado exclusivamente como uma medida da consistência da tendência histórica, e não como indicador de capacidade preditiva. Um R^2 elevado indica que o indicador seguiu uma trajetória relativamente linear ao longo dos anos analisados, enquanto um R^2 baixo sugere que fatores externos, variações locais ou políticas específicas causaram flutuações que não seguem um padrão linear simples.

4.3.4.3 Agregação de Resultados

Após calcular a regressão linear e o R^2 para cada município individualmente, os resultados foram agregados em diferentes níveis:

- **Nível municipal:** cada município possui seus próprios valores de mudança anual (β_1) e R^2 .
- **Nível estadual:** cálculo da média aritmética simples dos valores municipais dentro de cada estado.
- **Nível regional:** cálculo da média aritmética simples dos valores municipais dentro de cada região geográfica.
- **Nível nacional:** cálculo da média aritmética simples considerando todos os municípios.

Essa agregação hierárquica permite análises comparativas em diferentes escalas territoriais, identificando quais estados e regiões apresentaram tendências mais favoráveis.

4.3.5 Desenvolvimento de Visualizações

A visualização de dados foi empregada como técnica fundamental para comunicar os resultados das análises de forma clara, intuitiva e acessível. Para cada indicador, foram desenvolvidas visualizações que facilitassem a interpretação dos padrões temporais e espaciais identificados nas análises estatísticas.

4.3.5.1 Gráficos de Evolução Temporal com Tendência

Para representar visualmente a evolução de cada indicador ao longo do tempo, foram criados gráficos de linha que combinam:

- **Série temporal observada:** linha conectando os valores reais do indicador ano a ano.
- **Reta de tendência:** linha reta representando o modelo de regressão linear ajustado aos

dados.

- **Intervalo temporal:** eixo x representando os anos da série histórica.
- **Valor do indicador:** eixo y representando os valores do indicador em escala percentual.

Esses gráficos permitem visualizar simultaneamente os valores observados (com suas flutuações naturais) e a direção geral da mudança ao longo do tempo, facilitando a interpretação das tendências. O coeficiente de determinação (R^2) pode ser exibido como anotação no gráfico, fornecendo informação sobre a consistência da tendência linear.

4.3.5.2 *Gráficos de Comparação*

Foram desenvolvidos gráficos de barras para comparação entre regiões, estados ou períodos, apresentando métricas como média, valor máximo e valor mínimo dos indicadores. Esses gráficos facilitam a identificação visual de disparidades regionais e permitem comparações diretas entre diferentes unidades geográficas.

4.3.5.3 *Mapas Geográficos Interativos*

Para facilitar a visualização das disparidades regionais e da evolução espacial dos indicadores, foram elaborados mapas interativos do Brasil utilizando a biblioteca React Simple Maps. Os mapas foram desenvolvidos com as seguintes características:

- **Nível de agregação:** apresentação de dados agregados por estado (Unidade Federativa).
- **Escala de cores temática:** gradiente de cores que representa diferentes valores dos indicadores.
- **Interatividade:** funcionalidades de *hover* que exibem valores exatos ao passar o mouse sobre cada estado.
- **Filtros:** possibilidade de alternar entre diferentes indicadores e períodos temporais.

A visualização geográfica permite identificar rapidamente quais regiões e estados apresentaram valores mais críticos ou mais favoráveis nos indicadores analisados.

4.3.6 *Desenvolvimento do Dashboard Interativo*

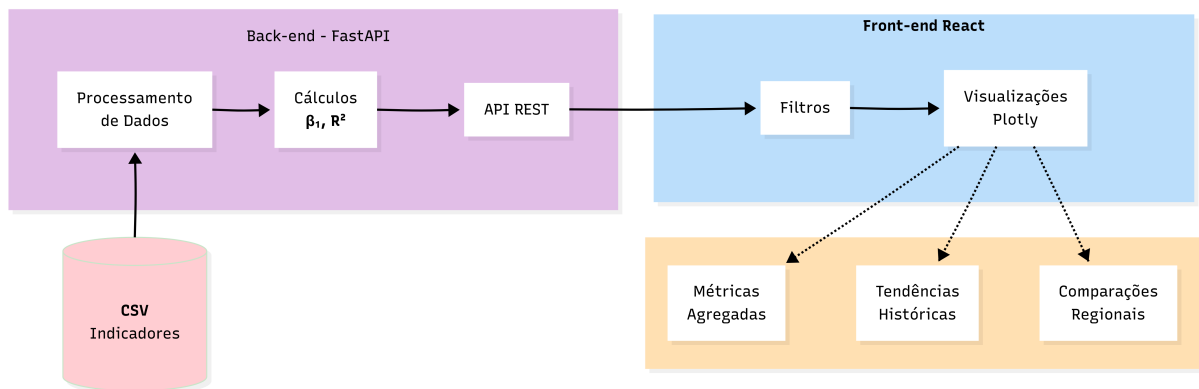
Após as análises exploratórias, cálculo de tendências e desenvolvimento das visualizações, iniciou-se a etapa de desenvolvimento do *dashboard* interativo. O *dashboard* foi projetado para consolidar todas as análises realizadas em uma interface web acessível, permitindo

que diferentes *stakeholders* do sistema educacional explorem os dados de forma intuitiva.

4.3.6.1 Arquitetura do Sistema

O *dashboard* foi desenvolvido seguindo uma arquitetura cliente-servidor que separa as responsabilidades entre backend e frontend, conforme ilustrado na Figura 2.

Figura 2 – Arquitetura do sistema do *dashboard*, mostrando a separação entre back-end (FastAPI), front-end (React) e camada de análises.



Fonte: Elaborado pela autora.

4.3.6.1.1 Back-end (Python com FastAPI)

O back-end é responsável pelo processamento dos dados educacionais e disponibilização via API REST. Foi desenvolvido utilizando FastAPI, um framework Python de alto desempenho para construção de APIs, que oferece validação automática de dados, documentação interativa e suporte a operações assíncronas (Ramírez, 2018). As principais responsabilidades do back-end incluem:

- Leitura e carregamento dos arquivos CSV pré-processados;
- Aplicação de filtros conforme requisições do frontend (região, estado, período);
- Cálculo de estatísticas descritivas (médias, desvios padrão, máximos, mínimos);
- Execução de regressões lineares e cálculo de R^2 para análise de tendências;
- Preparação e formatação de dados para visualizações;
- Exposição de endpoints REST que o frontend consome para obter dados específicos.

O back-end utiliza as bibliotecas Pandas para manipulação de dados, NumPy para operações numéricas e Scikit-learn para cálculo de métricas de regressão.

4.3.6.1.2 Front-end (React com TypeScript)

O front-end é responsável pela interface do usuário e apresentação das visualizações. Foi desenvolvido utilizando React 18.3 com TypeScript, Vite como ferramenta de *build*, e Tailwind CSS para estilização responsiva. As bibliotecas Plotly.js e Recharts foram utilizadas para gráficos interativos, e React Simple Maps para mapas geográficos. A biblioteca shadcn/ui forneceu componentes de interface reutilizáveis e acessíveis.

Essa separação de responsabilidades assegura que a camada de análise de dados (Python/back-end) seja especializada e otimizada para processamento eficiente, enquanto o React (front-end) se concentra na interface e na experiência do usuário. Como resultado, o sistema ganha em facilidade de manutenção, escalabilidade e permite futuras expansões.

4.3.6.2 Funcionalidades Implementadas

O *dashboard* oferece as seguintes funcionalidades principais:

- **Filtros interativos:** seleção de indicador (TDI, Abandono, Aprovação, Reprovação), região geográfica, estados específicos e intervalo temporal (*slider* para anos)
- **Visualizações em abas:**
 - Gráfico de tendência temporal com reta de regressão linear.
 - Gráfico de comparação entre regiões/estados (médias, máximos, mínimos).
 - Mapa interativo do Brasil com coloração temática por estado.
 - Tabela de dados detalhados com ordenação.
- **Cards de métricas:** exibição de total de registros, número de municípios, valores médios, máximos e mínimos dos indicadores.
- **Interatividade:** *hover* para visualizar valores exatos, zoom em gráficos e responsividade.

A funcionalidade principal do sistema é permitir a exploração dinâmica dos dados educacionais, facilitando a identificação de padrões temporais e disparidades regionais.

4.3.6.3 Disponibilidade e Acesso

O código-fonte completo do *dashboard* desenvolvido neste trabalho está disponível publicamente no repositório GitHub². O repositório contém tanto o código do back-end (Python/FastAPI) quanto do front-end (React/TypeScript), incluindo instruções detalhadas para

² <https://github.com/leticiaacx/educational-indicators-dashboard>

instalação e execução local do sistema.

Adicionalmente, uma versão demonstrativa do *dashboard* está hospedada e pode ser acessada diretamente pela plataforma Vercel³. Esta versão permite a exploração interativa dos dados educacionais sem necessidade de instalação local, facilitando o acesso por gestores educacionais, pesquisadores e demais interessados.

³ <https://educational-indicators-dashboard.vercel.app>

5 RESULTADOS

Este capítulo apresenta os resultados obtidos a partir das análises descritivas realizadas sobre os quatro indicadores educacionais selecionados: Taxa de Distorção Idade-Série (TDI), Taxa de Abandono, Taxa de Aprovação e Taxa de Reprovação. Os resultados são organizados por indicador, apresentando estatísticas nacionais, análises regionais e visualizações que facilitam a interpretação dos padrões identificados. A discussão dos resultados contextualiza os achados dentro do cenário educacional brasileiro, relacionando-os com políticas públicas e desafios persistentes.

5.1 Análise da Taxa de Distorção Idade-Série (TDI)

A Taxa de Distorção Idade-Série (TDI) é um indicador fundamental para avaliar a eficiência dos sistemas educacionais, pois mede a proporção de alunos com dois ou mais anos de atraso escolar em relação à série esperada para sua idade. Valores elevados de TDI indicam problemas estruturais como repetência sucessiva, ingresso tardio no sistema educacional ou interrupções na trajetória escolar.

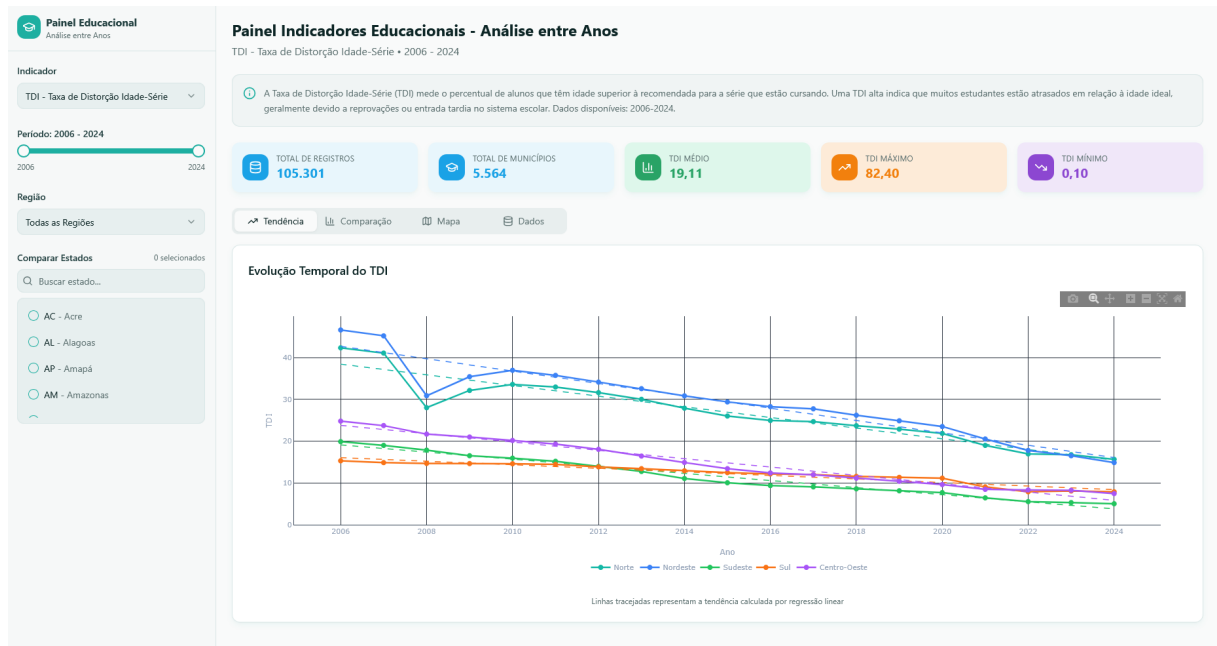
5.1.1 Resultados Nacionais

A análise longitudinal dos dados de TDI para o período de 2006 a 2024 revelou uma tendência decrescente consistente em nível nacional, conforme apresentado na interface do *dashboard* (Figura 3). Os principais resultados nacionais indicam uma melhoria substancial e sistemática ao longo de quase duas décadas.

O Brasil analisou dados de aproximadamente 5.564 municípios durante o período, em alguns casos variando o número de municípios, revelando uma mudança anual média de $-1,02$ pontos percentuais por ano. Essa redução consistente acumulou uma mudança total de $-18,41$ pontos percentuais entre 2006 e 2024, representando uma variação percentual de $-67,04\%$ no período completo. Em termos absolutos, o TDI médio nacional partiu de aproximadamente $28,5\%$ em 2006, alcançando $9,4\%$ em 2024.

Esses resultados sugerem que políticas educacionais voltadas para correção de fluxo escolar, combate à repetência e incentivo à permanência dos alunos tiveram impacto significativo em nível nacional. A magnitude da mudança anual média demonstra que, em média, a cada ano o Brasil conseguiu reduzir aproximadamente 1 ponto percentual do indicador, o que representa

Figura 3 – Interface do *dashboard* apresentando a evolução nacional do TDI (2006-2024) com linha de tendência e estatísticas descritivas.



Fonte: Elaborado pela autora.

um progresso gradual mas sustentado ao longo do período analisado.

5.1.2 Análise Regional

A análise regional dos dados do TDI revelou disparidades significativas nas tendências de melhoria entre as regiões do Brasil. A métrica principal utilizada foi a Mudança/Ano, obtida por regressão linear simples, que quantifica a variação média anual do índice TDI ao longo do período 2006–2024. Essa métrica representa a inclinação da reta de tendência, indicando a taxa média de melhoria (ou declínio) por ano em cada região.

Os resultados agregados por região geográfica são apresentados na Tabela 1, contendo: o número de municípios analisados, a taxa de mudança anual média, o coeficiente de determinação médio (R^2) e a variação percentual total no período de 2006 a 2024

Tabela 1 – Análise de tendências do TDI por região geográfica (2006–2024).

Região	N_Mun	Mudança/Ano	R^2	Var_%
Sul	1190	-0,427	0,929	-48,5%
Sudeste	1663	-0,865	0,978	-74,8%
Centro-Oeste	466	-1,007	0,977	-70,0%
Norte	450	-1,283	0,899	-63,0%
Nordeste	1793	-1,481	0,905	-68,1%

Fonte: Elaborado pela autora.

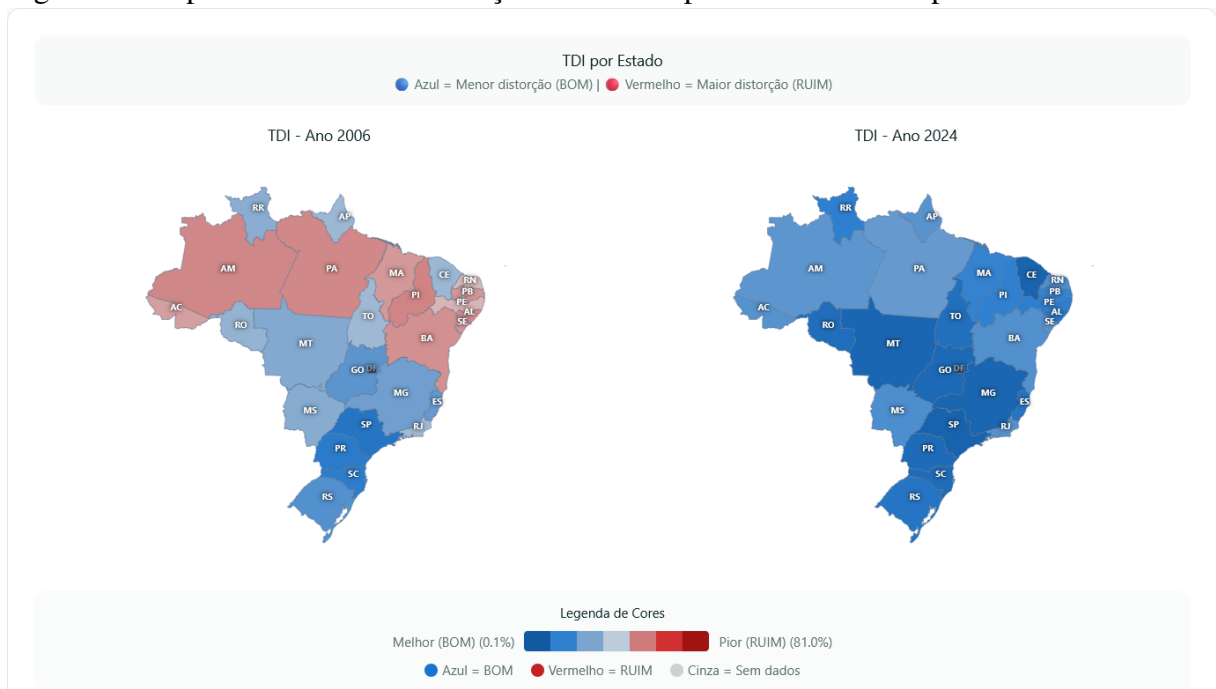
Os resultados regionais revelam padrões importantes:

5.1.2.1 Nordeste e Norte – Maiores Melhorias

As regiões Nordeste e Norte apresentaram as maiores taxas de redução no TDI durante o período analisado, como podemos observar na distribuição espacial apresentada na Figura 4. O Nordeste liderou com mudança anual de $-1,48$ pontos percentuais, seguido pela região Norte com $-1,28$ pontos percentuais anuais. Essas regiões, historicamente caracterizadas por maiores desafios educacionais e indicadores socioeconômicos mais desfavoráveis, demonstraram as melhorias mais expressivas.

A variação percentual total superior a 70% nessas regiões sugere que políticas nacionais de educação, como o Programa Nacional de Alimentação Escolar (PNAE), Bolsa Família (posteriormente Auxílio Brasil), programas de correção de fluxo e investimentos em infraestrutura escolar, podem ter tido maior impacto onde os desafios eram mais severos. O coeficiente de determinação médio (R^2) de 0,905 no Nordeste indica que 90,5% da variação observada no TDI ao longo do tempo pode ser explicada por uma tendência linear decrescente, demonstrando consistência e sistematicidade na melhoria do indicador.

Figura 4 – Mapa do Brasil com coloração temática representando o TDI por estado.

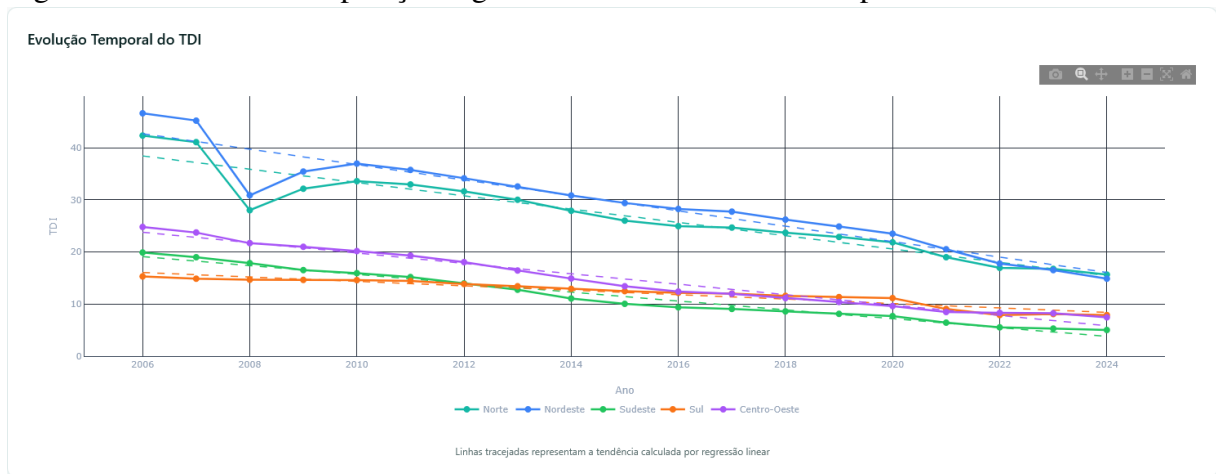


Fonte: Elaborado pela autora.

5.1.2.2 Sul, Centro-Oeste e Sudeste – Melhorias Moderadas

As regiões Sul, Centro-Oeste e Sudeste apresentaram reduções menos acentuadas no TDI, com mudanças anuais de $-0,42$, $-1,00$ e $-0,86$ pontos percentuais, respectivamente, como pode ser observado na comparação regional apresentada na Figura 5. A região Sul, com o menor declínio anual, apresentou o maior R^2 médio ($0,929$), indicando que apesar da menor magnitude de mudança, a tendência foi excepcionalmente consistente ao longo do período.

Figura 5 – Gráfico de comparação regional mostrando médias e dispersões do TDI.



Fonte: Elaborado pela autora.

Esses resultados podem ser explicados, em parte, pelo fato dessas regiões já partirem de valores iniciais de TDI inferiores à média nacional em 2006, apresentando menor margem para melhorias percentuais expressivas. O fenômeno observado é característico de processos de convergência regional, onde regiões com piores indicadores iniciais tendem a apresentar taxas de melhoria mais aceleradas.

5.1.2.3 Consistência das Tendências

Os valores de R^2 médio variaram entre $0,929$ (Sul) e $0,905$ (Nordeste), indicando que as tendências lineares decrescentes explicam bem o comportamento do TDI ao longo do tempo em todas as regiões. Valores de R^2 próximos a $0,90$ sugerem que a redução do TDI foi consistente e sistemática, não resultado de flutuações aleatórias. Essa robustez estatística reforça a interpretação de que as melhorias observadas refletem mudanças estruturais no sistema educacional brasileiro, resultantes de políticas públicas sustentadas ao longo de quase duas décadas.

5.2 Análise das Taxas de Rendimento Escolar

As taxas de rendimento escolar – Aprovação, Reprovação e Abandono – são componentes fundamentais para o cálculo do IDEB e refletem diretamente a capacidade dos sistemas educacionais em promover a progressão regular dos estudantes. A análise dessas taxas abrange o período de 2007 a 2024 (18 anos).

5.2.1 Taxa de Aprovação

5.2.1.1 Resultados Nacionais

A taxa de aprovação apresentou tendência crescente em nível nacional, com mudança anual média de +0,72 pontos percentuais por ano ($\beta_1 = +0,72$), obtida por regressão linear simples no período 2007–2024. Esse crescimento representa uma variação absoluta de 12,3 pontos percentuais e uma variação percentual de 15,5%.

Esses resultados indicam que o Brasil conseguiu aumentar consistentemente as taxas de aprovação, sugerindo melhorias tanto no desempenho acadêmico dos alunos quanto em políticas de avaliação e progressão escolar. O crescimento sustentado ao longo de 18 anos reflete o impacto positivo de programas de reforço escolar, políticas de progressão continuada e investimentos na qualificação docente.

5.2.1.2 Análise Regional

A análise regional das taxas de aprovação revelou um padrão consistente de convergência, com as regiões historicamente mais desiguais apresentando os maiores ganhos percentuais. O Nordeste liderou o crescimento com variação de 25,3% durante o período, apresentando R^2 médio de 0,88. Esse valor indica que 88% da variação observada nas taxas de aprovação ao longo dos 18 anos pode ser explicada por uma tendência linear crescente, demonstrando um padrão sistemático e consistente de melhoria.

A região Norte acompanhou essa tendência com crescimento de 20,6% e R^2 médio de 0,77, consolidando um padrão de convergência regional que caracteriza o período. O Centro-Oeste apresentou crescimento moderado de 12,0% com R^2 médio de 0,91, o mais alto entre todas as regiões, indicando linearidade excepcional na evolução das taxas de aprovação. Esse resultado sugere políticas educacionais implementadas de forma particularmente consistente ao

longo do tempo naquela região.

O Sudeste registrou variação de 10,7% com R^2 médio de 0,86, enquanto a região Sul, que já partia de taxas de aprovação historicamente mais elevadas, registrou o menor crescimento relativo (6,9%) com R^2 médio de 0,75. Esse padrão ilustra o efeito teto: regiões com indicadores iniciais já favoráveis têm menor margem para melhorias percentuais expressivas.

A Tabela 2 sintetiza os resultados regionais da taxa de aprovação.

Tabela 2 – Variação regional da taxa de aprovação (2007–2024).

Região	Variação (%)	R^2 médio
Nordeste	+25,3	0,88
Norte	+20,6	0,77
Centro-Oeste	+12,0	0,91
Sudeste	+10,7	0,86
Sul	+6,9	0,75

Fonte: Elaborado pela autora.

5.2.2 Taxa de Reprovação

5.2.2.1 Resultados Nacionais

A taxa de reprovação apresentou tendência decrescente em nível nacional, com mudança anual média de $-0,45$ pp/ano ($\beta_1 = -0,45$), reduzindo de 13,8% (2007) para 6,1% (2024). Essa redução representa uma variação absoluta de $-7,7$ pontos percentuais e uma variação percentual de $-55,8\%$, com mudança anual média de $-0,45$ pontos percentuais por ano.

A redução de mais da metade das taxas de reprovação ao longo de 18 anos representa um avanço significativo na promoção do fluxo escolar adequado. Esse resultado está associado à implementação de políticas de progressão continuada, programas de reforço escolar e mudanças nos critérios de avaliação que privilegiam o desenvolvimento progressivo das competências dos estudantes.

5.2.2.2 Análise Regional

A análise regional das taxas de reprovação revelou reduções substanciais em todas as regiões do país. O Nordeste apresentou o maior declínio relativo de 69,5% no período, com R^2 médio de 0,779. Essa redução expressiva reflete políticas de progressão continuada e programas

de reforço escolar que tiveram maior impacto nas regiões com maiores desafios históricos.

A região Norte registrou redução de 62,4% com R^2 médio de 0,674, acompanhando o padrão de convergência observado nos demais indicadores. Surpreendentemente, as regiões Sudeste e Centro-Oeste apresentaram as maiores reduções percentuais (79,4% e 76,0%, respectivamente), com R^2 médios elevados (0,854 e 0,891). Esses valores altos de R^2 indicam que as reduções nessas regiões foram altamente lineares e consistentes ao longo do período.

É importante contextualizar que essas regiões partiam de taxas iniciais de reprovação inferiores, o que explica as reduções percentuais mais expressivas quando os valores absolutos já são baixos. A região Sul apresentou redução de 62,6% com R^2 médio de 0,732, indicando boa consistência na tendência de redução.

A Tabela 3 apresenta a síntese dos resultados regionais da taxa de reprovação.

Tabela 3 – Variação regional da taxa de reprovação (2007–2024).

Região	Variação (%)	R^2 médio
Sudeste	–79,4	0,854
Centro-Oeste	–76,0	0,891
Nordeste	–69,5	0,779
Sul	–62,6	0,732
Norte	–62,4	0,674

Fonte: Elaborado pela autora.

5.2.3 Taxa de Abandono

5.2.3.1 Resultados Nacionais

A taxa de abandono escolar apresentou forte tendência decrescente em nível nacional, mudança anual média de $-0,27$ pp/ano ($\beta_1 = -0,27$), de 6,7% para 2,1%. Essa redução representa uma variação absoluta de $-4,6$ pontos percentuais e uma variação percentual de $-68,7\%$.

A redução de quase 70% nas taxas de abandono representa uma das conquistas mais expressivas do período analisado, refletindo políticas de combate à evasão escolar e programas de transferência de renda condicionada à frequência escolar, como o Bolsa Família. A consistência dessa redução ao longo de 18 anos demonstra o impacto estrutural dessas políticas na permanência dos estudantes no sistema educacional.

5.2.3.2 Análise Regional

A análise regional das taxas de abandono revelou reduções excepcionais em todas as regiões do país. As regiões Nordeste e Norte apresentaram as maiores reduções, com 85,7% e 81,8% de declínio no período, respectivamente. O Nordeste destacou-se com R^2 médio de 0,930, o mais elevado entre todos os indicadores analisados, indicando que 93% da variação observada nas taxas de abandono pode ser explicada por uma tendência linear decrescente altamente consistente.

A região Norte apresentou R^2 médio de 0,857, também indicando alta consistência na tendência de declínio. O Centro-Oeste registrou a maior redução percentual entre todas as regiões (95,0%) com R^2 médio de 0,866, enquanto o Sudeste apresentou redução de 89,8% com R^2 médio de 0,827.

A região Sul, embora tenha apresentado a menor redução percentual (68,3%), exibiu o maior R^2 médio (0,941), indicando que a redução do abandono foi excepcionalmente linear e consistente ao longo dos 18 anos analisados. Valores de R^2 superiores a 0,90 são raros em análises de séries temporais de dados educacionais, reforçando a robustez e a consistência das tendências observadas.

A Tabela 4 apresenta a síntese dos resultados regionais da taxa de abandono.

Tabela 4 – Variação regional da taxa de abandono (2007–2024).

Região	Variação (%)	R^2 médio
Centro-Oeste	–95,0	0,866
Sudeste	–89,8	0,827
Norte	–81,8	0,857
Nordeste	–85,7	0,930
Sul	–68,3	0,941

Fonte: Elaborado pela autora.

5.3 Síntese dos Principais Insights

A análise descritiva dos quatro indicadores educacionais ao longo das séries históricas de 2006–2024 (TDI) e 2007–2024 (Abandono, Aprovação, Reprovação) revelou padrões consistentes e significativos.

5.3.1 Tendências Positivas Generalizadas

Todos os quatro indicadores analisados apresentaram tendências favoráveis no período: TDI com redução de 67%, Taxa de Abandono com redução de aproximadamente 69%, Taxa de Reprovação com redução de 56%, e Taxa de Aprovação com aumento de 15,5%.

Essas melhorias sugerem impacto positivo de políticas educacionais nacionais implementadas ao longo das últimas duas décadas, incluindo programas de correção de fluxo, transferência de renda condicionada, alimentação escolar e investimentos em infraestrutura.

5.3.2 Convergência Regional

As regiões Norte e Nordeste, historicamente mais desiguais em termos educacionais, apresentaram as maiores melhorias relativas em todos os quatro indicadores. Esse padrão indica um processo de convergência regional, com redução gradual das disparidades históricas entre regiões mais e menos desenvolvidas.

A convergência regional é particularmente evidente quando se observa que as taxas de melhoria no Nordeste e Norte foram consistentemente superiores às demais regiões, os valores de R^2 elevados indicam que essas melhorias foram sistemáticas, não episódicas, e a diferença entre as regiões com melhores e piores indicadores diminuiu ao longo do período.

5.3.3 Padrões Identificados nos Indicadores de Fluxo Escolar

As análises descritivas dos indicadores de fluxo escolar revelaram *insights* importantes sobre as tendências regionais no período de 2007 a 2024. Na taxa de aprovação, as regiões Nordeste e Norte apresentaram os maiores aumentos durante o período, com crescimento consistente em todas as regiões e redução gradual das disparidades regionais.

Na taxa de reprovação, Nordeste e Norte apresentaram as maiores reduções, com declínio consistente nas taxas em âmbito nacional, embora algumas regiões tenham mantido taxas relativamente estáveis em determinados períodos. Na taxa de abandono, Norte e Nordeste apresentaram as maiores reduções, com forte declínio em todas as regiões, embora desafios persistentes permaneçam em algumas áreas específicas.

Esses resultados revelam que as regiões Norte e Nordeste, historicamente mais desiguais em termos educacionais, apresentaram as maiores melhorias relativas no período analisado, sugerindo impacto positivo de políticas nacionais de educação voltadas para redução

de desigualdades regionais.

5.3.4 *Consistência das Tendências*

Os valores de R^2 consistentemente elevados (acima de 0,77 na maioria dos casos, chegando a 0,94 em algumas análises regionais) indicam que as tendências identificadas são robustas e sistemáticas. Não se trata de flutuações aleatórias, mas sim de mudanças estruturais consistentes ao longo de quase duas décadas.

5.3.5 *Desafios Persistentes*

Apesar das melhorias significativas, alguns desafios permanecem: o TDI médio nacional em 2024 ainda é de aproximadamente 9,4%, indicando que cerca de 1 em cada 10 alunos ainda está com dois ou mais anos de atraso escolar; disparidades regionais, embora reduzidas, ainda são significativas; alguns estados específicos apresentam indicadores substancialmente piores que a média nacional; e a necessidade de políticas direcionadas persiste para áreas com progresso mais lento.

6 CONCLUSÃO

Este trabalho teve como objetivo desenvolver um *dashboard* interativo para análise e identificação de tendências em indicadores educacionais brasileiros, utilizando dados históricos disponibilizados pelo INEP. A motivação central foi transformar grandes volumes de dados educacionais em visualizações intuitivas e insights acionáveis que pudessem apoiar gestores, pesquisadores e formuladores de políticas públicas na compreensão da evolução da qualidade educacional no Brasil.

6.1 Principais Resultados

As análises descritivas realizadas sobre séries históricas de 2006 a 2024 (TDI) e 2007 a 2024 (demais indicadores) revelaram resultados significativos:

6.1.1 Tendências Positivas Generalizadas

Todos os quatro indicadores apresentaram melhorias consistentes e estatisticamente ao longo do período analisado. O TDI apresentou redução de 67,04%, passando de aproximadamente 28,5% em 2006 para 9,4% em 2024. A taxa de abandono escolar reduziu 68,7%, a taxa de reprovação caiu 55,8%, e a taxa de aprovação aumentou 15,5% no período. Essas melhorias sugerem impacto positivo de políticas educacionais nacionais implementadas ao longo das últimas duas décadas.

6.1.1.1 Consistência das Tendências

Os valores de R^2 consistentemente elevados (acima de 0,80 na maioria das análises regionais) indicam que as tendências identificadas não resultam de flutuações aleatórias, mas sim de mudanças estruturais sistemáticas ao longo de quase duas décadas. A consistência das tendências reforça a confiabilidade dos resultados e sugere que os avanços observados refletem transformações reais no sistema educacional brasileiro.

6.1.2 Contribuições do Trabalho

Este trabalho oferece contribuições práticas, metodológicas e teóricas para o campo da análise de dados educacionais:

6.1.2.1 *Contribuição Prática*

O *dashboard* interativo desenvolvido constitui uma ferramenta prática que centraliza a visualização de séries históricas extensas de indicadores educacionais, permitindo que gestores e pesquisadores explorem tendências temporais e comparações regionais de forma intuitiva.

A arquitetura cliente-servidor desenvolvida, com backend em Python/FastAPI e frontend em React/TypeScript, demonstra a viabilidade de integrar análises estatísticas robustas com interfaces modernas e responsivas. O sistema pode ser expandido futuramente para incluir novos indicadores, análises adicionais ou funcionalidades mais avançadas.

6.1.2.2 *Contribuição Metodológica*

Este trabalho demonstra a aplicação de técnicas de regressão linear para análise descritiva de tendências históricas, diferenciando-se de abordagens preditivas. A utilização do coeficiente de determinação (R^2) como medida da consistência de tendências passadas, e não como indicador de capacidade preditiva, oferece uma perspectiva metodológica clara sobre análise exploratória de séries temporais educacionais.

O processo metodológico detalhado – desde a coleta automatizada de dados até o desenvolvimento do *dashboard* – pode servir como referência para trabalhos futuros que busquem desenvolver ferramentas similares para análise de dados educacionais ou de outras áreas de políticas públicas.

6.1.2.3 *Contribuição Teórica*

Os resultados empíricos obtidos contribuem para a literatura sobre evolução dos indicadores educacionais brasileiros, fornecendo evidências quantitativas sobre tendências de longo prazo e convergência regional. A identificação de que as regiões Norte e Nordeste apresentaram as maiores melhorias relativas reforça a importância de políticas nacionais direcionadas à redução de desigualdades.

6.1.3 *Limitações do Estudo*

É importante reconhecer as limitações deste trabalho para contextualizar adequadamente seus resultados e contribuições:

A principal limitação foi a elevada taxa de dados faltantes em 10 dos 14 indicadores

inicialmente considerados, o que inviabilizou análises mais abrangentes e motivou a reorientação dos objetivos. A análise ficou restrita aos quatro indicadores com séries históricas mais completas, impossibilitando a investigação de outros fatores potencialmente relevantes como adequação da formação docente, complexidade de gestão escolar e nível socioeconômico.

As análises realizadas são fundamentalmente descritivas e não estabelecem relações causais entre políticas específicas e as melhorias observadas.

O *dashboard* desenvolvido, embora funcional e intuitivo, apresenta funcionalidades básicas de filtragem e visualização. Análises mais sofisticadas, como comparações estatísticas formais entre regiões, testes de hipóteses ou modelagem de relações entre variáveis, não foram implementadas.

6.1.4 Trabalhos Futuros

Os resultados e limitações identificados neste trabalho apontam diversas direções promissoras para pesquisas futuras:

Futuras pesquisas podem buscar integrar dados de outras fontes (como Censo Escolar, PNAD Contínua, dados socioeconômicos municipais) para complementar os indicadores do INEP e possibilitar análises mais abrangentes. A inclusão de variáveis contextuais como PIB per capita municipal, IDH, investimento público em educação e perfil demográfico poderia enriquecer a compreensão dos fatores associados às melhorias observadas.

Com dados mais completos, trabalhos futuros podem aplicar técnicas estatísticas mais avançadas, como modelos de regressão múltipla, análise de painel, modelos multinível ou análise de mediação, para investigar relações causais entre políticas específicas e resultados educacionais. Técnicas de aprendizado de máquina, como *random forests* ou *gradient boosting*, podem ser exploradas para identificar os preditores mais importantes da qualidade educacional.

6.2 Considerações Finais

Este trabalho demonstrou que a análise descritiva sistemática de séries históricas extensas pode revelar padrões importantes sobre a evolução da qualidade educacional no Brasil. Os resultados obtidos indicam melhorias consistentes e generalizadas nos quatro indicadores analisados, com destaque para o processo de convergência regional que reduziu disparidades históricas.

Embora desafios educacionais persistam – com aproximadamente 9,4% dos alunos ainda em situação de distorção idade-série em 2024 e disparidades regionais significativas –, as tendências positivas identificadas ao longo de quase duas décadas oferecem evidências encorajadoras sobre o impacto de políticas educacionais nacionais. A continuidade desses esforços, aliada ao monitoramento sistemático de indicadores e à adoção de abordagens baseadas em evidências, é fundamental para a consolidação dos avanços e para a superação dos desafios remanescentes.

Por fim, este trabalho reforça a importância da disponibilização de dados abertos governamentais e do desenvolvimento de ferramentas que facilitem sua análise e interpretação. A transparência de informações educacionais e o acesso facilitado a análises fundamentadas são pilares essenciais para uma gestão pública mais eficiente, participativa e comprometida com a melhoria contínua da qualidade da educação brasileira.

REFERÊNCIAS

- Asimov Academy. **Regressão Linear: Conceitos fundamentais e aplicações**. 2025. Disponível em: <https://hub.asimov.academy/blog/regressao-linear/>. Acesso em: 8 jan. 2026.
- BOSCARDIN, C. K.; FERGUS, K. B.; HELLEVIG, B.; HAUER, K. E. Twelve tips to promote successful development of a learner performance dashboard within a medical education program. **Medical Teacher**, v. 40, n. 8, p. 855–861, 2018.
- BRASIL. **Lei nº 12.527, de 18 de novembro de 2011**: Regula o acesso a informações previstos no inciso xxxiii do art. 5º, no inciso ii do § 3º do art. 37 e no § 2º do art. 216 da constituição federal. Brasília, DF, 2011. Diário Oficial da União. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 13 jan. 2026.
- BROWN, M. J. **Applied Research Methods for Business and Management**. [S. l.]: Goodfellow Publishers, 2015. ISBN 978-1-910158-75-5.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 9. ed. São Paulo: Saraiva, 2017.
- CASTRO, L. N. de; FERRARI, D. G. **Introdução à Mineração de Dados: Conceitos básicos, algoritmos e aplicações**. São Paulo: Saraiva, 2017. ISBN 978-85-472-0781-8.
- CORTINA, R. C. **Desenvolvimento e Validação de Dashboards Educacionais em Sistemas de Gestão Educacional**. Dissertação (Dissertação (Mestrado em Tecnologias da Informação e Comunicação)) – Universidade Federal de Santa Catarina, Araranguá, 2025. Disponível em: <https://repositorio.ufsc.br/handle/123456789/265560>. Acesso em: 8 jan. 2026.
- COSTA, F. R. B.; SILVA, P. C. d. L. e.; GUIMARÃES, F. G.; BATISTA, L. S. Regressão linear aplicada na predição de séries temporais fuzzy. In: **Anais do XIII Simpósio Brasileiro de Automação Inteligente**. Porto Alegre: SBA, 2017. Disponível em: https://www.sba.org.br/Proceedings/SBAI/SBAI2017/SBAI17/papers/paper_99.pdf. Acesso em: 8 jan. 2026.
- DIO. **React**: Biblioteca JavaScript para interfaces de usuário. 2025. Disponível em: <https://www.dio.me>. Acesso em: 12 jan. 2026.
- ELLAWAY, R. H.; PUSIC, M. V.; GALBRAITH, R. M.; CAMERON, T. Developing the role of big data and analytics in health professions education. **Medical Teacher**, v. 36, n. 3, p. 216–222, 2014.
- FEW, S. **Information Dashboard Design: The effective visual communication of data**. Sebastopol, CA: O’Reilly Media, 2006. ISBN 978-0-596-10016-2.
- FONSECA, S. O. d.; NAMEN, A. A. Mineração em bases de dados do INEP: Uma análise exploratória para nortear melhorias no sistema educacional brasileiro. **Educação em Revista**, v. 32, n. 1, p. 133–157, 2016.
- FRIENDLY, M.; WAINER, H. **A History of Data Visualization and Graphic Communication**. [S. l.]: Harvard University Press, 2021. ISBN 978-0-674-97527-5.

- GRINSTEIN, G.; WARD, M. O. Introduction to data visualization. In: FAYYAD, U.; GRINSTEIN, G. G.; WIERSE, A. (Ed.). **Information Visualization in Data Mining and Knowledge Discovery**. San Francisco, CA: Morgan Kaufmann, 2002. p. 21–45. ISBN 978-1-55860-689-6.
- HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Análise Multivariada de Dados**. 6. ed. Porto Alegre: Bookman, 2009.
- IBGE. **Estimativas da População**. 2024. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html>. Acesso em: 12 jan. 2026.
- INEP. **Indicadores Educacionais**. 2024. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/indicadores-educacionais>. Acesso em: 12 jan. 2026.
- Jeduca. **Análise de Indicadores Educacionais por Região**. 2023. Disponível em: <https://jeduca.org.br>. Acesso em: 12 jan. 2026.
- Leme Tec. **Dashboard e Análise de Dados Educacionais**. 2024. Disponível em: <https://www.leme.tec.br/dashboard>. Acesso em: 8 jan. 2026.
- MCKINNEY, W. Data structures for statistical computing in Python. In: WALT, S. van der; MILLMAN, J. (Ed.). **Proceedings of the 9th Python in Science Conference**. [S. l.: s. n.], 2010. p. 51–56.
- MENEZES, M. **Dashboard Interativo para Análise de Indicadores Educacionais: Uma abordagem baseada em dados abertos**. Dissertação (Dissertação (Mestrado em Informática)) – Universidade Federal do Ceará, Fortaleza, 2023.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. 5. ed. Hoboken, NJ: John Wiley & Sons, 2012. ISBN 978-0-470-54281-1.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2. ed. São Paulo: Editora Edgar Blucher, 2006.
- Observatório de Educação. **Plataformas de Dados Educacionais**. 2016. Disponível em: <https://observatoriodeeducacao.institutounibanco.org.br>. Acesso em: 12 jan. 2026.
- Observatório de Educação. **Dados Abertos em Educação**. 2021. Disponível em: <https://observatoriodeeducacao.institutounibanco.org.br>. Acesso em: 12 jan. 2026.
- Python Software Foundation. **Python Language Reference, Version 3.12**. 2024. Disponível em: <https://docs.python.org/3/>. Acesso em: 12 jan. 2026.
- QEdu. **Portal QEdu: Dados educacionais**. 2024. Disponível em: <https://www.qedu.org.br>. Acesso em: 12 jan. 2026.
- QUEMY, A. Data pipeline selection and optimization. In: **Proceedings of the... (Nome da Conferência)**. [S. l.: s. n.], 2019.
- RAJ, A.; BOSCH, J.; OLSSON, H. H.; WANG, T. J. Modelling data pipelines. In: **2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)**. [S. l.: s. n.], 2020. p. 13–20.

RAMÍREZ, S. **FastAPI Framework**. 2018. Disponível em: <https://fastapi.tiangolo.com>. Acesso em: 12 jan. 2026.

REIS, M. M.; MENEZES, M. **Análise de Séries Temporais**. Florianópolis: Universidade Federal de Santa Catarina, 2017. Disponível em: <https://www.inf.ufsc.br/~marcelo.menezes.reis/Cap4.pdf>. Acesso em: 8 jan. 2026.

Revisa. **Métodos Estatísticos para Análise de Tendências Educacionais**. 2025. Disponível em: <https://revisa.com.br>. Acesso em: 12 jan. 2026.

ROSSUM, G. V.; DRAKE, F. L. **Python 3 Reference Manual**. Scotts Valley, CA: CreateSpace, 2009. ISBN 978-1-4414-1269-0.

THESARI, S. S.; TROJAN, F.; BATISTUS, D. R. A decision model for municipal resources management. **Management Decision**, v. 57, n. 11, p. 3015–3034, 2019.

TUKEY, J. W. **Exploratory Data Analysis**. Reading, MA: Addison-Wesley, 1977. ISBN 978-0-201-07216-8.

UFPR. **Coeficiente de Determinação R^2** . Curitiba: Laboratório de Estatística e Geoinformação, 2024. Disponível em: <http://leg.ufpr.br/~lucambio/CE313/20241S/Regressao2.pdf>. Acesso em: 8 jan. 2026.

UNWIN, A. Why is data visualization important? What is important in data visualization? **Harvard Data Science Review**, v. 2, n. 1, p. 1–7, 2020.

VALLE-CRUZ, D.; FERNANDEZ-CORTEZ, V.; GIL-GARCIA, J. R. From E-budgeting to smart budgeting: Exploring the potential of artificial intelligence in government decision-making for resource allocation. **Government Information Quarterly**, v. 39, n. 2, p. 101644, 2022.