



**INSTITUTO
FEDERAL**
Brasília

Instituto Federal de Educação, Ciência e Tecnologia de Brasília, campus Taguatinga,
Campus Taguatinga

O USO DOS DADOS DO SAEB COMO FERRAMENTA DE GESTÃO ESCOLAR NO DF

Por

BRUNO GOMES DA SILVA, FERNANDO MIGUEL ALCÂNTARA DE OLIVEIRA

Trabalho de Graduação

BRASÍLIA/2024

Bruno Gomes da Silva, Fernando Miguel Alcântara de Oliveira

**O USO DOS DADOS DO SAEB COMO FERRAMENTA DE GESTÃO
ESCOLAR NO DF**

Trabalho apresentado ao Curso de Ciência da Computação
do Instituto Federal de Educação, Ciência e Tecnologia de
Brasília, *campus* Taguatinga, como requisito parcial para
obtenção do grau de Bacharel

Orientador: Raimundo Cláudio da Silva Vasconcelos

BRASÍLIA
2024

Ficha de identificação da obra elaborada pelo bibliotecário
Marcelo José Rodrigues da Conceição (CRB1-2323)

Silva, Bruno Gomes da

S586u O uso dos dados do SAEB como ferramenta de gestão escolar no DF / Bruno Gomes da Silva, Fernando Miguel Alcântara de Oliveira. Brasília-DF, 2024.

44 f. : il.

Trabalho de Conclusão de Curso de Bacharelado em Ciência da Computação, Instituto Federal de Educação, Ciência e Tecnologia de Brasília, Campus Taguatinga, 2024.

Orientador: Raimundo Cláudio da Silva Vasconcelos.

Inclui referências.

1. Organização escolar - Distrito Federal (Brasil). 2. Ensino médio - Distrito Federal (Brasil). 3. Análise de dados. 4. Banco de dados da Web. I. Título. II. Oliveira, Fernando Miguel Alcântara de. III. Vasconcelos, Raimundo Cláudio da Silva. IV. Instituto Federal de Educação, Ciência e Tecnologia de Brasília.

CDU 37.07:004(817.4)

Trabalho de Graduação apresentada por **Bruno Gomes da Silva, Fernando Miguel Alcântara de Oliveira** ao programa de Graduação em Ciência da Computação do Instituto Federal de Brasília, sob o título **O Uso dos Dados do SAEB Como Ferramenta de Gestão Escolar no DF**, orientada pelo **Raimundo Cláudio da Silva Vasconcelos** e aprovada pela banca examinadora formada pelos professores:

Prof. Ms. Diego Martins de Oliveira
Instituto Federal de Brasília/Campus Taguatinga

Prof. Ms Thiago Batista Amorim
Instituto Federal de Brasília/Campus Taguatinga

Prof. Dr. Raimundo Cláudio da Silva Vasconcelos
Instituto Federal de Brasília/Campus Taguatinga

Agradecimentos

Hoje, neste momento de celebração e gratidão, gostaríamos de expressar nosso profundo agradecimento por todo o apoio, orientação e inspiração que recebemos ao longo de nossa jornada acadêmica e, em particular, durante a realização do nosso trabalho de conclusão de curso.

Raimundo, você tem sido um guia excepcional durante toda essa jornada. Sua dedicação incansável ao ensino, sua paixão pelo conhecimento e seu compromisso com o sucesso de seus alunos são verdadeiramente inspiradores. Sua paciência, disponibilidade para ouvir nossas dúvidas e orientação precisa foram fundamentais para o nosso crescimento acadêmico e profissional. Agradecemos por terem acreditado em nós, por desafiarem-nos constantemente a irmos além dos limites e por nos incentivarem a explorar novos horizontes.

Aos nossos queridos colegas da turma da Sociedade da Uva, somos gratos por todo o apoio mútuo, trocas enriquecedoras e amizades que se desenvolveram ao longo desses anos. Cada um de vocês trouxe habilidades únicas, perspectivas diversas e uma incrível energia coletiva para nossa turma. Juntos, enfrentamos desafios, superamos obstáculos e compartilhamos momentos inesquecíveis. O trabalho em equipe que desenvolvemos foi fundamental para o nosso crescimento individual e para o sucesso de todos.

Além disso, gostaríamos de estender nossa gratidão a todos os professores, instrutores e profissionais que contribuíram direta ou indiretamente para nossa formação acadêmica. Seus ensinamentos, exemplos e experiências compartilhadas moldaram nosso caminho, estimularam nossa curiosidade intelectual e nos capacitaram para enfrentar os desafios futuros.

Finalmente, expressamos nossa profunda gratidão às nossas famílias e amigos que estiveram ao nosso lado durante toda essa jornada. Seu apoio incondicional, incentivo constante e amor inabalável foram o alicerce que nos sustentou nos momentos de dificuldade e nos inspirou a perseverar em busca de nossos sonhos.

À medida que concluimos esta etapa de nossas vidas acadêmicas, olhamos para trás com orgulho e para o futuro com entusiasmo. Cada uma das pessoas mencionadas neste texto desempenhou um papel crucial em nossa jornada e somos profundamente gratos por cada uma delas.

Que este espírito de gratidão e reconhecimento nos acompanhe ao longo de nossas vidas, inspirando-nos a sempre valorizar e apreciar aqueles que contribuem para nosso crescimento e sucesso.

Com sinceridade e gratidão,
Bruno Silva e Fernando Oliveira.

Resumo

O Sistema de Avaliação da Educação Básica (SAEB) é uma ferramenta essencial para avaliar a qualidade do ensino fundamental e médio no Brasil. Este trabalho tem como problema a falta de ferramentas de análise que relacionem o desempenho dos estudantes nas provas do SAEB com as condições socioeconômicas obtidas por meio de questionários. A hipótese é que a disponibilização de painéis de *Business Intelligence* (BI), que cruzem esses dados, pode fornecer *insights* valiosos para a gestão escolar. A pesquisa, de caráter aplicado, explicativo e experimental, desenvolveu e implementou esses painéis em um sistema web, utilizando dados de alunos do Distrito Federal, com o objetivo de permitir uma visualização mais clara e detalhada, ajudando gestores a tomar decisões mais informadas e estratégicas. Conclui-se que os questionários socioeconômicos aplicados aos alunos são essenciais para entender o desempenho dos estudantes, demonstrando que fatores como escolaridade dos pais, renda familiar e infraestrutura doméstica podem ter influência direta nos resultados das provas, auxiliando na formulação de políticas educacionais mais eficazes.

Palavras-chave: SAEB, Educação Básica, Análise de Dados, DF, *Data Warehouse*, Sistema *Web*, Gestão Educacional.

Abstract

The Basic Education Evaluation System (SAEB) is an essential tool for assessing the quality of elementary and secondary education in Brazil. This study addresses the lack of analytical tools that relate student performance in SAEB tests to the socioeconomic conditions gathered through questionnaires. The hypothesis is that the availability of Business Intelligence (BI) dashboards, which cross-reference these data, can provide valuable insights for school management. The research, which is applied, explanatory, and experimental in nature, developed and implemented these dashboards in a web-based system, using data from students in the Federal District. The goal was to enable clearer and more detailed visualization, helping school administrators make more informed and strategic decisions. The study concludes that the socioeconomic questionnaires applied to students are essential for understanding their performance, showing that factors such as parents' education, family income, and household infrastructure can have a direct influence on test results, thus supporting the formulation of more effective educational policies.

Keywords: SAEB, Basic Education, Data Analysis, Federal District, Data Warehouse, Web System, Educational Management.

Lista de Figuras

2.1	Exemplo de parte do questionário socioeconômico aplicado aos alunos	13
2.2	Funcionamento do sistema de uma <i>Data Warehouse</i>	15
2.3	Processo de criação de um trabalho de análise de dados	16
3.1	Distribuição dos alunos conforme classe econômica para escolas públicas e privadas	18
3.2	Desempenho do 3º ano do Ensino Médio em relação à proficiência em Matemática.	20
3.3	Estágios na implementação do framework de BI	21
3.4	Proficiência média dos estudantes de escolas públicas em língua portuguesa e matemática - 2005 à 2013	22
3.5	Proficiência média dos estudantes de escolas privadas em língua portuguesa e matemática - 2005 à 2013	22
4.1	Microdados do SAEB	28
4.2	Fluxo de Processamento e Visualização dos Dados do SAEB	29
5.1	Exemplo de uso do Painel do 5º ano	32
5.2	Exemplo de uso do painel do 9º ano	33
5.3	Exemplo de uso do painel do 3º ano do ensino médio	33
5.4	Filtro de Perguntas e Respostas	34
5.5	Exemplo porcentagem da resposta selecionada	34
5.6	Painel de escolas	35
5.7	Exemplo de uso do painel de escolas	35
5.8	Filtro Nível socioeconômico	36
5.9	Filtro de Ano, Localização e Tipo de Escola	36
5.10	Exemplo de Porcentagem dos Níveis Socioeconômico	36
5.11	Exemplos de informações contidas no Painel de Escolas	36
5.12	Exemplos de Medianas de acordo com o Ano de 2013 a 2021	37

Lista de Tabelas

3.1	Termos chave para levantamento bibliográfico	17
3.2	Dados dos 8 artigos levantados	26

Sumário

1	Introdução	9
1.1	Organização do trabalho	10
1.2	Justificativa	10
1.3	Objetivos	10
1.3.1	Objetivo geral	10
1.3.2	Objetivos específicos	11
2	Contextualização	12
2.1	SAEB	12
2.2	<i>Data Warehouse</i>	13
2.3	<i>Business Intelligence</i> (BI)	14
2.4	Ferramentas	15
3	Levantamento Bibliográfico	17
3.1	Trabalhos correlatos	17
3.1.1	Fatores associados ao desempenho de alunos do 3º ano do ensino médio do Distrito Federal : uma análise multinível aplicada a dados do Saeb 2013	18
3.1.2	Fatores Associados ao Desempenho escolar em Matemática do SAEB 2011 para alunos do 5º ano no Distrito Federal	19
3.1.3	O que os dados do SAEB nos dizem sobre o desempenho dos estudantes em Matemática?	19
3.1.4	Uma proposta de aplicação de <i>Business Intelligence</i> no sistema educacional brasileiro	20
3.1.5	DESEMPENHO ESCOLAR E ESCOLHA ENTRE ESCOLAS PÚBLICAS E PRIVADAS NO RIO GRANDE DO SUL: EVIDÊNCIAS EMPÍRICAS A PARTIR DE DADOS DO SAEB	21
3.1.6	A CONSTRUÇÃO DE UM <i>DATA WAREHOUSE</i> UTILIZANDO OS INDICADORES EDUCACIONAIS DO INEP	22
3.1.7	Descoberta de conhecimento através da análise e mineração em dados do ENEM	23
3.1.8	Uso de <i>Business Intelligence</i> para avaliação de indicadores de desempenho na educação básica: um estudo de caso no estado do Acre	24
3.1.9	Comparativo entre os trabalhos correlatos	24
4	Desenvolvimento	27
4.1	Metodologia	27
4.2	Coleta e Análise de Dados	28

5 Resultados	32
5.1 Limitações da Pesquisa	37
6 Conclusões	38
Referências	40

1

Introdução

O sistema educacional brasileiro reflete a distribuição de responsabilidades entre municípios, estados e governo federal. É estabelecido que os municípios, estados e o Distrito Federal atuem na educação infantil, ensino fundamental e médio. Além disso, o governo federal desempenha um papel redistributivo e supletivo, oferecendo assistência técnica e financeira. Adicionalmente, além da educação regular, há também a educação superior. Essa estrutura complexa visa abranger as diversas necessidades educacionais da sociedade brasileira (MENEZES, 2001).

Em 1990, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), uma autarquia federal vinculada ao Ministério da Educação, instituiu o Sistema de Avaliação da Educação Básica (SAEB). O propósito central desse sistema é conduzir uma análise diagnóstica da educação básica no Brasil, gerando dados que possam orientar a formulação, ajuste e monitoramento de políticas públicas nos âmbitos municipal, estadual e federal. O intuito é contribuir para aprimorar a qualidade, equidade e eficiência do ensino (INEP, 2013).

Atualmente, o SAEB ocorre bianualmente, abrangendo estudantes do ensino fundamental e do ensino médio. Essas avaliações fornecem dados detalhados sobre o desempenho dos alunos em disciplinas fundamentais, como Língua Portuguesa e Matemática. Adicionalmente, o SAEB disponibiliza informações abrangentes sobre escolas, professores e resultados de questionários socioeconômicos, oferecendo uma visão ampla e detalhada do contexto educacional.

Dessa forma, a realização de uma análise mais aprofundada dos dados do SAEB, principalmente em relação aos questionários socioeconômicos pode auxiliar na gestão escolar do Distrito Federal. Pois essas informações seriam fundamentais para a elaboração de um plano de ação voltado para as aprendizagens dos estudantes, possibilitando, por exemplo, a implementação de abordagens mais direcionadas às atividades de reforço escolar.

Atualmente, os dados disponibilizados pelo SAEB não correlacionam informações socioeconômicas com o desempenho dos alunos. Portanto, decorre a necessidade de conduzir uma análise destinada a avaliar se há alguma correlação as respostas dos questionários socioeconômicos com o desempenho dos alunos no DF.

1.1 Organização do trabalho

O trabalho está dividido da seguinte forma: o Capítulo 1 é composto pela proposta, justificativa, objetivo geral e objetivos específicos. O Capítulo 2 apresenta a contextualização, fundamentação teórica, descrevendo o Sistema de Avaliação da Educação Básica (SAEB) e as tecnologias que foram utilizadas, no capítulo 3 são apresentados os trabalhos correlatos. Os demais aspectos relacionados ao desenvolvimento estão detalhados no capítulo 4. Em seguida pode-se visualizar os resultados no Capítulo 5. Por fim, o capítulo 6 descreve as conclusões do projeto.

1.2 Justificativa

A análise dos dados do Sistema de Avaliação da Educação Básica (SAEB) do Distrito Federal busca apresentar relevância significativa por várias razões:

- Busca proporcionar uma compreensão mais detalhada do desempenho educacional no DF, levando em consideração a intrincada estrutura do sistema educacional brasileiro;
- Permitir a identificação e a investigação de possíveis disparidades educacionais entre alunos, visando entender as causas subjacentes;
- Possibilita uma fácil navegação pelas informações favorecendo a gestão escolar viabilizando identificar padrões, tendências e eventos que influenciam o desempenho educacional no DF;
- Buscar não apenas gerar conhecimento, mas também contribuir para a tomada de decisões informadas no âmbito educacional, promovendo melhorias efetivas com base em evidências;

1.3 Objetivos

A seguir são explanados o objetivo geral e os objetivos específicos deste trabalho.

1.3.1 Objetivo geral

Neste trabalho, pretende-se realizar análises dos dados disponibilizados pelo SAEB no período de 2013 a 2021 com foco no Distrito Federal. Além disso, pretende-se criar painéis de BI que possibilitem a gestores educacionais uma compreensão mais detalhada da qualidade educacional no DF. Para isso serão disponibilizados painéis em um Sistema *Web* que correlacionam o desempenho dos estudantes com os questionários socioeconômicos.

1.3.2 Objetivos específicos

- Obter, transformar e carregar dados do SAEB entre 2013 e 2021;
- Analisar e validar os dados;
- Possibilitar a correlação de informações socioeconômicas com o desempenho dos alunos;
- Construir painéis de BI;
- Disponibilizar os resultados em um Sistema *Web*;

2

Contextualização

Para uma melhor compreensão do foco desse trabalho, serão apresentados o SAEB, as tecnologias e por fim as ferramentas utilizadas neste trabalho.

2.1 SAEB

O Sistema de Avaliação da Educação Básica (SAEB, 2023), conduzido pelo INEP, exerce um papel essencial na avaliação da qualidade da educação básica no Brasil. Este conjunto de avaliações, realizado a cada dois anos, abrange tanto a rede pública quanto uma amostra da rede privada, utilizando testes e questionários para refletir sobre os níveis de aprendizagem dos estudantes. Ao analisar esses resultados no contexto de outras informações relevantes, o SAEB proporciona às escolas e redes municipais e estaduais uma ferramenta valiosa para avaliar a qualidade da educação, além de fornecer subsídios para a formulação e aprimoramento de políticas educacionais baseadas em evidências.

Iniciado em 1990 o SAEB, passou por aprimoramentos ao longo das edições, incorporando testes específicos de língua portuguesa e matemática em diferentes etapas do ensino fundamental e médio. Os testes de língua portuguesa focam na leitura, avaliando a capacidade dos alunos de compreender textos em diversos níveis. Paralelamente, os testes de matemática enfatizam a resolução de problemas, considerando habilidades como observação, estabelecimento de relações e argumentação. Atualmente, secretários, diretores e professores respondem eletronicamente questionários que abordam funcionamento das redes de ensino, informações sobre gestão, atividades e infraestrutura das escolas, formação, experiência, condições de trabalho e práticas pedagógicas.

Em relação aos alunos, além dos testes cognitivos, eles respondem a um questionário impresso que visa levantar temas como a condição socioeconômica, participação da família, interação com professores e colegas em sala de aula, além de atividades pedagógicas. A Figura 2.1 ilustra alguns itens desse questionário aplicado aos alunos do 9º ano do Ensino Fundamental no ano de 2021.

Figura 2.1: Exemplo de parte do questionário socioeconômico aplicado aos alunos

ID Censo: 99999 **Sistema de Avaliação da Educação Básica - SAEB 2021**
Cód. Aluno: 99999 **QUESTIONÁRIO DO(A) ALUNO(A) - 9º ANO DO ENSINO FUNDAMENTAL**

Questão 15. Você se utiliza de transporte escolar, ou passe escolar, para ir à escola?

A Não.
 B Sim.

Questão 16. Com que idade você entrou na escola?

A 3 anos ou menos.
 B 4 ou 5 anos.
 C 6 ou 7 anos.
 D 8 anos ou mais.

Questão 17. A partir do primeiro ano do ensino fundamental, em que tipo de escola você estudou?

A Somente em escola pública.
 B Somente em escola particular.
 C Em escola pública e em escola particular.

Questão 18. Você já foi reprovado(a)?

A Não.
 B Sim, uma vez.
 C Sim, duas vezes ou mais.

Questão 19. Alguma vez você abandonou a escola deixando de frequentá-la até o final do ano escolar?

A Nunca.
 B Sim, uma vez.
 C Sim, duas vezes ou mais.

Questão 20. Fora da escola e em dias de aula, quanto tempo você usa para: (Preencha um círculo em cada linha)

	Não uso meu tempo para isso	Menos de 1 hora	Entre 1 e 2 horas	Mais de 2 horas
a) Estudar (lição de casa, trabalhos escolares, etc.).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b) Fazer cursos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c) Trabalhar em casa (lavar louça, limpar quintal, cuidar dos irmãos).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d) Trabalhar fora de casa (recebendo ou não um salário).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e) Lazer (TV, internet, brincar, música etc.).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fonte: (SAEB, 2020)

Além do desempenho cognitivo, a qualidade da educação envolve as complexas interações entre o ambiente escolar e as relações estabelecidas entre os diversos sujeitos envolvidos no processo educacional. O SAEB, ao incorporar essa abordagem abrangente, não apenas fornece uma avaliação dos conhecimentos adquiridos, mas também oferece *insights* fundamentais para uma compreensão mais holística do cenário educacional brasileiro.

2.2 Data Warehouse

Um *Data Warehouse* (Cetax, 2020) é uma plataforma destinada à armazenagem e centralização de dados e informações relacionados a um tema específico. Ele tem a capacidade de organizar e consolidar dados provenientes de bancos de dados, facilitando a análise de grandes volumes de informações e fornecendo dados que são úteis para embasar a tomada de decisões.

Como fundamento para qualquer sistema de dados, o *Data Warehouse* viabiliza a exploração eficiente desses dados, graças à centralização abrangente de todas as informações. Isso resulta em maior eficácia na aquisição e aplicação desses dados. A inclusão de dados em um *Data Warehouse* oferece uma abordagem para superar os desafios relacionados à falta de uniformidade dos dados, uma questão evidenciada neste estudo devido à diferença temporal

entre os dados analisados. Após serem processados e normalizados por meio de ferramentas ou linguagens de programação, esses dados podem ser reutilizados e atualizados de maneira mais ágil e segura, constituindo uma base sólida para futuros trabalhos.

A Figura 2.2 ilustra o funcionamento do sistema em que as fontes de dados 1, 2 e 3 são processadas através do método de extração, transformação e carga (ETL). O ETL é um processo essencial no gerenciamento de dados, que envolve três etapas principais: a extração (onde os dados são coletados de diferentes fontes), a transformação (em que os dados são limpos, formatados e convertidos para um formato adequado ao objetivo da análise), e a carga (onde os dados transformados são inseridos no data warehouse). Em seguida, os dados são encaminhados para o *data warehouse*, onde ocorrem as etapas subsequentes de visualização, geração de relatórios e aplicação de inteligência empresarial.

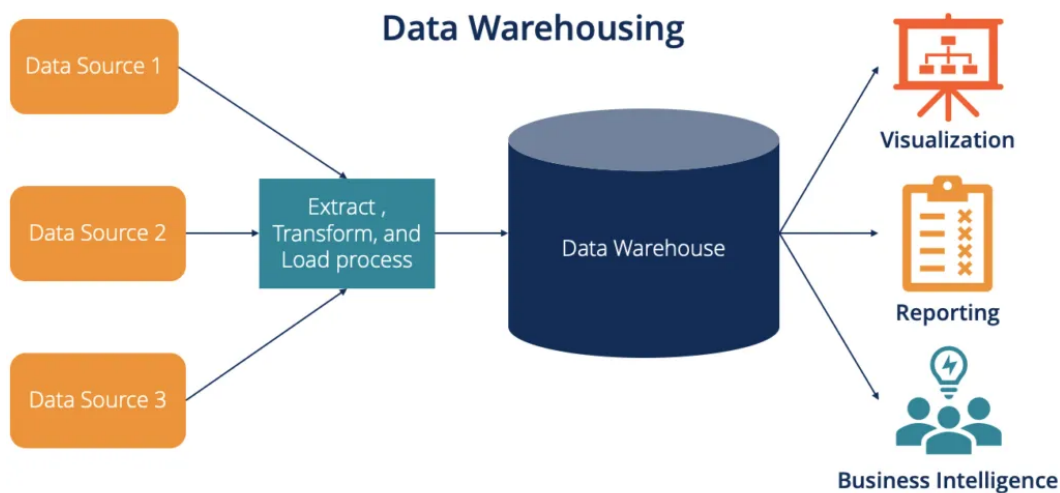
2.3 Business Intelligence (BI)

Conforme (GILAD; GILAD, 1985), o termo *Business Intelligence* (BI) começou a ser empregado para descrever uma ferramenta de coleta e análise de dados, com o propósito de apoiar o processo de tomada de decisões. A partir desse conceito, o BI evoluiu para fornecer suporte computacional aos gestores e à alta administração das organizações, abrangendo recursos como a geração de relatórios dinâmicos e multidimensionais, análises de tendências, previsões e identificação de fatores críticos para o sucesso organizacional (TURBAN et al., 2009).

De maneira similar, (HABUL; PILAV-VELIC, 2010) apontam que o BI é composto por etapas como a coleta, tratamento e análise de dados, além do alinhamento dessas informações estratégicas às necessidades das organizações. A implementação de sistemas baseados no BI é fundamental para promover o crescimento eficaz das organizações, uma vez que auxilia no desenvolvimento de estratégias e na construção de uma base sólida de conhecimento, permitindo uma melhor definição dos próximos passos, redução de falhas e maximização do desempenho.

Além disso, o BI pode ser compreendido como um processo de extração, transformação, gestão e análise de dados, utilizando diversos frameworks ou protocolos, com o objetivo de gerar informações e conhecimentos valiosos (FITRIANA; ERIYATNO; DJATNA, 2011). Isso possibilita sua aplicação prática por meio de diferentes métodos e ferramentas, sempre levando em conta o contexto específico de cada organização, uma vez que cada ambiente possui características próprias e exige abordagens adequadas.

A análise dos dados do SAEB foi baseada em BI pois é uma abordagem mais ampla que engloba o processo de coleta, transformação, análise e distribuição de dados estratégicos para apoiar a gestão organizacional. O BI oferece um método estruturado para analisar grandes volumes de dados e identificar tendências, prever cenários futuros e avaliar fatores críticos para o sucesso. Essa abordagem não só facilita a formulação de estratégias mais embasadas, como também ajuda a minimizar riscos e maximizar o desempenho organizacional, tornando-se um elemento chave para empresas que buscam aprimorar seus processos de tomada de decisão.

Figura 2.2: Funcionamento do sistema de uma *Data Warehouse*

Fonte: (Corporate Finance Institute, 2024)

2.4 Ferramentas

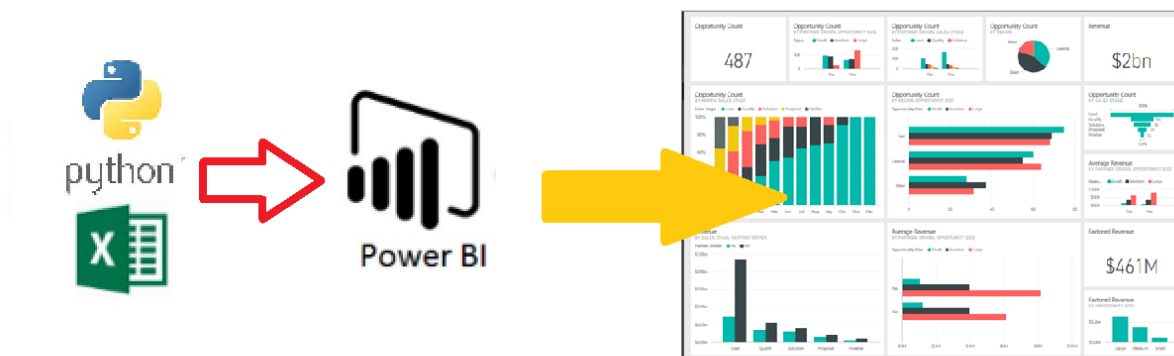
Para este trabalho utilizaram-se as seguintes tecnologias:

- **Microsoft Power BI** (Microsoft, 2023a) é um *software* lançado em 2015 que oferece recursos avançados de visualização e *business intelligence*. Este programa possui a capacidade de carregar visualizações personalizadas por meio do *AppSource*, possibilita o acesso a informações em tempo real, está disponível em aplicativos móveis, é voltado para uso empresarial, recebe atualizações mensais com novas funcionalidades e é conhecido por sua facilidade de uso.

Devido à sua excelente integração com as mais diversas fontes de dados, bem como à extensa variedade de opções gráficas nativas do programa e a participação ativa dos usuários na criação de novas formas de apresentação de dados, o Power BI foi empregado na elaboração de um *dashboard* utilizando dados do SAEB. Essa abordagem permite que os usuários que acessam a plataforma obtenham as respostas desejadas em relação aos dados selecionados.

A Figura 2.3 apresenta como o *Power BI* interage com os demais produtos *Microsoft* e com a linguagem *Python*.

Figura 2.3: Processo de criação de um trabalho de análise de dados



Fonte: Autor

- **Excel** (Microsoft, 2023b), foi lançado em 1987, é uma aplicação para criação de planilhas eletrônicas. Muitas empresas dependem atualmente de suas bases de dados operacionais nesses formatos de planilha. O Excel destaca-se por sua interface intuitiva, excelentes ferramentas de cálculo e construção de gráficos, fatores que, aliados a uma estratégia de marketing agressiva, o tornaram um dos aplicativos de computador mais populares desenvolvidos até hoje.

Ele foi selecionado para agilizar o trabalho, aproveitando sua facilidade para a compreensão de parte dos dados, principalmente os dicionários das bases de dados fornecidos pelo SAEB. A planilha eletrônica Excel ofereceu uma excelente plataforma para criar tabelas e visualização.

- **Python**, uma linguagem de programação (Python Software Foundation, 2023), foi concebida em 1992, embora tenha ganhado notoriedade mais recentemente por sua sintaxe relativamente simples e fácil compreensão. Seu crescimento em popularidade ocorreu principalmente entre profissionais na indústria tecnológica, incluindo engenheiros, matemáticos, cientistas de dados, pesquisadores e outros. *Python* é uma linguagem de alto nível, projetada para priorizar a legibilidade em detrimento da velocidade do código, e geralmente requer menos linhas de código quando comparada a programas similares em outras linguagens.

Neste projeto, *Python* foi a linguagem escolhida para extração, transformação e carregamento dos dados. Pois, possui grande capacidade de lidar com conjuntos de dados extensos e é suportada por bibliotecas robustas voltadas para análise de dados e estatística, como *Numpy* (para computações científicas) e *Pandas* (para manipulação e análise de dados).

3

Levantamento Bibliográfico

3.1 Trabalhos correlatos

Com o objetivo de definir o escopo deste projeto, foram examinados e analisados estudos correlatos, buscando fundamentar as escolhas de ferramentas para o desenvolvimento do protótipo, bem como aqueles que abordaram o SAEB. As pesquisas relacionadas são detalhadas a seguir.

Os termos apresentados na tabela 3.1 foram combinados de diversas formas e utilizados para pesquisa em diversas ferramentas de busca disponíveis na *internet*. Dessa forma os artigos apresentados nas seções a seguir foram encontrados através do *Google Scholar*. Os principais trabalhos avaliados estão detalhados a seguir.

Português (PT-BR)	Inglês (EN)
SAEB	<i>SAEB</i>
Análise de Dados	<i>Data analysis</i>
Inteligência empresarial	<i>Business Intelligence</i>
Armazém de Dados	<i>Data Warehouse</i>
Educação Básica	<i>Basic education</i>
Distrito Federal	<i>Federal District</i>
Desempenho escolar	<i>School Performance</i>

Tabela 3.1: Termos chave para levantamento bibliográfico

Este trabalho busca analisar os dados do SAEB e sua aplicação na gestão educacional das escolas do Distrito Federal, de forma semelhante ao que foi realizado no artigo de (JUNIOR, 2021), em que foi proposta a aplicação de Business Intelligence (BI) no sistema educacional brasileiro utilizando dados do INEP e IDEB. Assim como no estudo mencionado, onde o BI foi utilizado para melhorar a tomada de decisões no contexto educacional, este trabalho também visa criar painéis de BI para apoiar a gestão escolar, permitindo uma análise detalhada das informações.

Além disso, seguindo uma abordagem similar à de (ILHA et al., 2021), que desenvolveu um data warehouse com indicadores educacionais do INEP para as escolas de Santa Maria, RS, este trabalho também faz uso de técnicas de ETL para organizar e transformar os dados do SAEB

em um repositório centralizado, facilitando a visualização e a análise dos resultados. Enquanto o estudo de Ilha et al. focou na validação da eficácia dos painéis de controle em melhorar a gestão escolar, este trabalho se concentra na criação de uma ferramenta de BI acessível e prática para apoiar os gestores do Distrito Federal na identificação de áreas críticas e na tomada de decisões estratégicas utilizando questionários socioeconômicos do SAEB aplicados aos alunos.

3.1.1 Fatores associados ao desempenho de alunos do 3º ano do ensino médio do Distrito Federal : uma análise multinível aplicada a dados do Saeb 2013

O estudo (BATISTA, 2016), investigou-se o desempenho dos alunos do 3º ano do Ensino Médio no Distrito Federal, utilizando dados da Avaliação Nacional da Educação Básica (Aneb) de 2013. Esta avaliação, integrante do Sistema de Avaliação da Educação Básica (SAEB), compreende provas de matemática e língua portuguesa.

Foram empregadas ferramentas de Análise Multinível, que é um caso específico da regressão usado para estudar populações estruturadas de forma hierárquica, e regressão multinível, com a linguagem de programação R, utilizada para lidar com questões estatísticas durante a preparação e análise dos dados. Os resultados revelaram que, na escala de proficiência do SAEB, o desempenho médio em língua portuguesa situou-se em 282,5, correspondendo ao nível 3 da escala, enquanto em matemática, a média foi de 292, também dentro da faixa do nível 3 da escala de proficiência. Estes dados fornecem uma visão detalhada do desempenho dos alunos, permitindo uma compreensão mais aprofundada dos fatores que influenciam seu rendimento acadêmico. No artigo, também foi examinada a distribuição dos alunos de acordo com a classe econômica em escolas públicas e privadas. Esta distribuição está apresentada na Figura 3.1.

Figura 3.1: Distribuição dos alunos conforme classe econômica para escolas públicas e privadas

Classe	Pública	Privada
A1	4 (0,2%)	60 (5,5%)
A2	108 (4,6%)	363 (33,1%)
B1	369 (16,0%)	373 (34,0%)
B2	718 (31,1%)	215 (19,6%)
C1	595 (25,7%)	56 (5,1%)
C2	371 (16,1%)	19 (1,7%)
D	134 (5,8%)	4 (0,4%)
Sem Resposta	12 (0,5%)	8 (0,7%)
Total	2311 (100,0%)	1098 (100,0%)

Fonte: (BATISTA, 2016)

Em síntese, o estudo destaca a relevância de variáveis específicas na explicação do desempenho dos alunos do 3º ano do Ensino Médio no Distrito Federal, conforme evidenciado pela abordagem multinível. A análise aponta que, além das diferenças no desempenho entre as disciplinas de matemática e língua portuguesa, fatores como o sexo dos alunos e o tipo de dependência administrativa da escola desempenham papéis distintos na explicação dessas variações.

3.1.2 Fatores Associados ao Desempenho escolar em Matemática do SAEB 2011 para alunos do 5º ano no Distrito Federal

O artigo (LIMA AMBROSIO, 2014), investiga-se os fatores determinantes do desempenho dos alunos do 5º ano do Distrito Federal na prova de Matemática do SAEB em 2011. Utilizando a técnica de análise multinível, com o aluno e a escola como níveis distintos, identifica-se que características específicas influenciam significativamente os resultados.

A pesquisa também utilizou o Censo Escolar 2011, para coletar informações sobre a infraestrutura escolar. A Análise Multinível foi a técnica aplicada, com a linguagem *R* sendo empregada na preparação e análise estatística dos dados.

Os resultados destacaram que no nível do aluno, observou-se que o sexo masculino e a ausência de reprovações foram os fatores mais relevantes para um desempenho satisfatório. Já no âmbito escolar, a média da nota da escola destacou-se como o fator mais significativo.

3.1.3 O que os dados do SAEB nos dizem sobre o desempenho dos estudantes em Matemática?

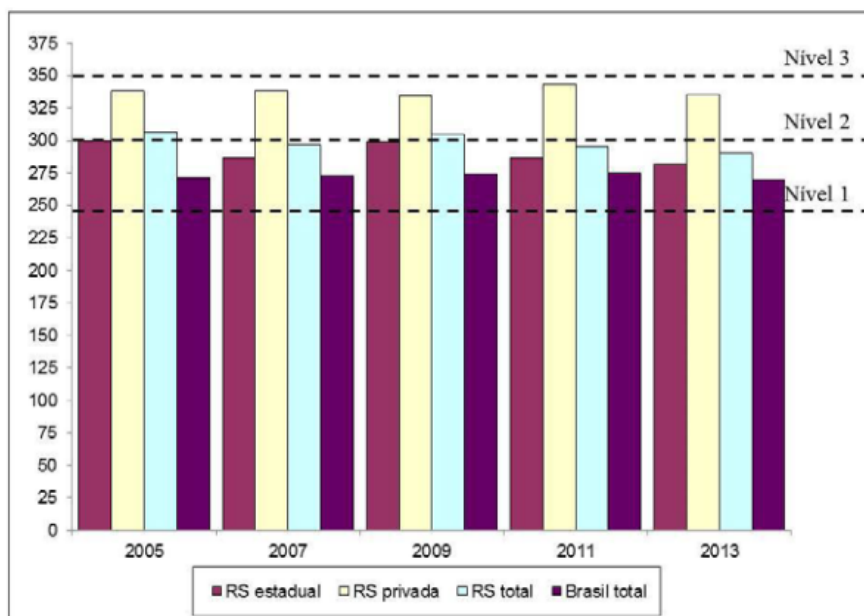
O estudo (PEREIRA; TOLENTINO-NETO, 2015), são abordadas as notas médias dos alunos do 5º e 9º anos do Ensino Fundamental e do 3º ano do Ensino Médio no Rio Grande do Sul em Matemática, nas edições de 2005 a 2013 do SAEB. A pesquisa busca compreender as oscilações e tendências nas médias de desempenho ao longo desses anos, com ênfase nas habilidades matemáticas associadas a cada nível de desempenho. Os resultados evidenciaram oscilações progressivas, avanços notáveis em diferentes momentos escolares e um déficit entre as habilidades esperadas e obtidas pelos alunos em todos os anos avaliados.

Os dados do SAEB, disponíveis no Portal do INEP, foram fundamentais para esta análise, fornecendo informações sobre o desempenho em diferentes esferas, como escolas, municípios, estados e regiões. A metodologia incluiu a utilização de notas numéricas de proficiência, agrupadas em níveis para oferecer um panorama das habilidades matemáticas desenvolvidas e aquelas que demandam maior atenção. A análise considerou escolas estaduais e privadas no Rio Grande do Sul, bem como médias totais do estado e do Brasil, permitindo inferir quais conteúdos matemáticos estão mais ou menos desenvolvidos em cada etapa escolar.

Os resultados revelaram um aumento progressivo nas médias do 5º ano, destacando a constante oscilação nas médias do 9º ano ao longo das edições analisadas. No 3º ano do

Ensino Médio, os alunos apresentaram lacunas em habilidades essenciais, como a operação com o plano cartesiano, identificação de intervalos e resolução de problemas envolvendo figuras espaciais. Este estudo oferece uma visão abrangente do desempenho matemático dos alunos no Rio Grande do Sul, fornecendo *insights* cruciais para aprimorar a educação nesta área. Na Figura 3.2 observa-se o desempenho do 3º ano do Ensino Médio em relação à proficiência em Matemática no Rio Grande do Sul.

Figura 3.2: Desempenho do 3º ano do Ensino Médio em relação à proficiência em Matemática.



Fonte: (PEREIRA; TOLENTINO-NETO, 2015)

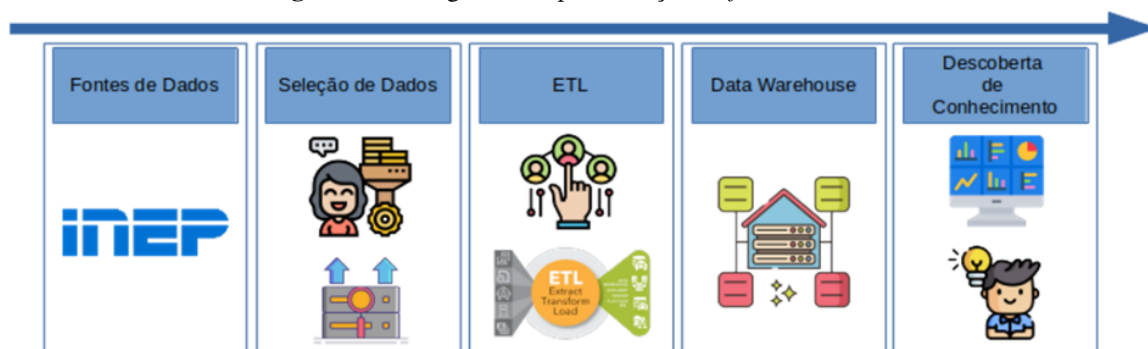
3.1.4 Uma proposta de aplicação de *Business Intelligence* no sistema educacional brasileiro

O estudo (JUNIOR, 2021) apresenta uma proposta inovadora ao aplicar *Business Intelligence* (BI), nos dados do sistema educacional brasileiro, disponibilizados pelo INEP e IDEB. BI é um conjunto de estratégias e ferramentas utilizadas para a coleta, análise e apresentação de informações de negócios, visando a tomada de decisões informadas. A pesquisa busca responder à questão sobre as contribuições e desafios associados ao uso de BI em dados educacionais no contexto brasileiro.

A metodologia adotada foi um estudo aplicado com objetivos exploratórios e descritivos, utilizando a ferramenta *Power BI* para implementar estágios de análise de dados. Estes estágios incluem a seleção de dados, ETL (Extração, Transformação e Carga), construção de um *data warehouse*, e descoberta de conhecimento. O processo de ETL é crucial para mover e transformar dados de várias fontes para um repositório centralizado, assegurando que os dados estejam limpos e prontos para análise. A ferramenta desenvolvida demonstrou resultados satisfatórios, gerando visualizações interativas e gráficas que facilitam a divulgação de informações e a tomada de

decisões, além de abrir espaço para técnicas avançadas como a mineração de dados. Mineração de dados refere-se ao processo de explorar grandes conjuntos de dados para identificar padrões, tendências e relações ocultas, utilizando algoritmos e técnicas estatísticas avançadas. Os estágios na implementação de um *framework* de *Business Intelligence* podem ser visualizados na Figura 3.3.

Figura 3.3: Estágios na implementação do *framework* de BI



Fonte: (JUNIOR, 2021)

Os dados fornecidos pelo INEP sobre a educação brasileira revelaram uma riqueza de informações, e a aplicação do BI se mostrou adequada como ferramenta de apoio para análises, decisões e implementações de ações no sistema educacional. A interface desenvolvida possibilita a fácil observação de dados gerais e específicos da educação, proporcionando uma análise mais aprofundada e compreensão dos fatos. Essa abordagem intuitiva do BI destaca-se como uma contribuição relevante para o acesso e interpretação de dados educacionais, promovendo uma melhor compreensão do cenário educacional brasileiro.

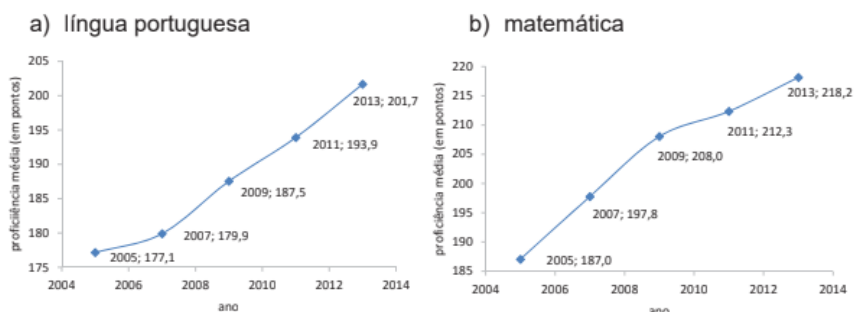
3.1.5 DESEMPENHO ESCOLAR E ESCOLHA ENTRE ESCOLAS PÚBLICAS E PRIVADAS NO RIO GRANDE DO SUL: EVIDÊNCIAS EMPÍRICAS A PARTIR DE DADOS DO SAEB

O artigo (PRESTES, 2016) tem por objetivo analisar os fatores que afetam a escolha entre a rede de ensino pública e privada pela família do aluno, além disso comparar o desempenho entre elas. Para isso, a autora utilizou dados do SAEB com foco no estado do Rio Grande do Sul e, inicialmente apontou que embora o estado apresentasse desempenho superior à média brasileira fora um dos estados que menos evoluiu em termos de proficiência média geral, que é o desempenho apresentado pelas escolas públicas e privadas conjuntamente, no período de 2005 à 2011.

Para o mesmo período, os alunos das escolas públicas e privadas do Rio Grande do Sul obtiveram um aumento nas médias de desempenho em língua portuguesa e matemática conforme evidenciado nas figuras 3.4 e 3.5.

Figura 3.4: Proficiência média dos estudantes em língua portuguesa e matemática - 2005 à 2013

Figura 1: Rio Grande do Sul – Proficiência média dos estudantes da 4ª série/5º ano do ensino fundamental de escolas públicas em língua portuguesa e matemática- 2005 a 2013

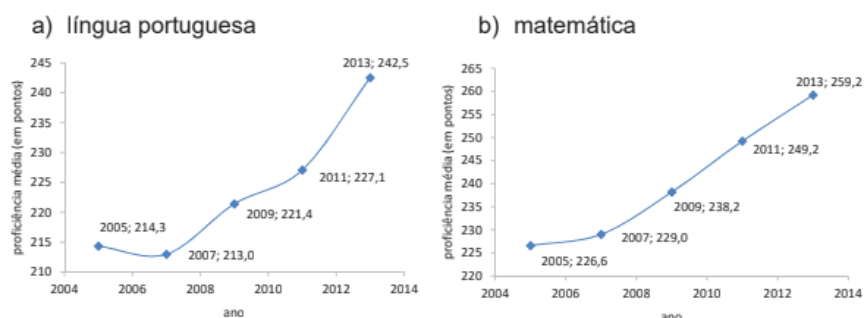


Fonte: Elaboração própria a partir de dados do SAEB de 2005 a 2013.

Fonte: (PRESTES, 2016)

Figura 3.5: Proficiência média dos estudantes em língua portuguesa e matemática - 2005 à 2013

Figura 3: Rio Grande do Sul – Proficiência média dos estudantes da 4ª série/5º ano do ensino fundamental de escolas privadas em língua portuguesa e matemática - 2005 a 2013



Fonte: Elaboração própria a partir de dados do SAEB de 2005 a 2013.

Fonte: (PRESTES, 2016)

Mesmo com essa melhora no desempenho, o artigo aponta que o Rio Grande do Sul não foi capaz de auferir destaque no comparativo com outros estados, além disso notou-se melhor desempenho dos alunos das escolas privadas. Outros fatores como estrutura escolar, frequência na pré-escola, o maior nível social e a escolaridade superior da mãe estão relacionados com uma maior proficiência escolar.

3.1.6 A CONSTRUÇÃO DE UM *DATA WAREHOUSE* UTILIZANDO OS INDICADORES EDUCACIONAIS DO INEP

O estudo (ILHA et al., 2021) foi conduzido nas escolas municipais de Santa Maria, RS, empregando indicadores educacionais do INEP, tais como: média de alunos por turma; horas-aula diárias; taxas de distorção idade-série; O autor detalhou o processo de carga de dados no *Power*

BI, enfatizando a transformação das variáveis. Em colaboração com a secretaria de educação local, o estudo validou a eficácia dos painéis de controle, revelando resultados favoráveis:

- 71,4% acreditavam em melhorias nos processos de gestão escolar.
- 75% previam uma nova visão do ambiente escolar.
- 67,8% reconheceram a utilidade na visualização dos ambientes escolares para políticas públicas.
- 78,6% recomendaram o produto para outros gestores.

A pesquisa buscou responder sobre a contribuição dos indicadores do INEP para a gestão educacional, propondo uma metodologia para importar e visualizar esses dados de forma acessível, visando melhorar o ensino-aprendizagem. Os objetivos foram alcançados, incluindo técnicas de modelagem de data warehouse, carregamento e visualização dos indicadores, uso de ferramenta de BI e análise da sua usabilidade. Desafios surgiram na interação com os participantes via e-mail devido à pandemia, exigindo criatividade e persistência para estimular a resposta quantitativa.

A pesquisa evidenciou a necessidade de valorizar dados brutos na gestão escolar, destacando o caso do Google, que baseia decisões em dados, não apenas opiniões dos gestores. Essa valorização, apesar dos desafios enfrentados na interação com os usuários, mostrou-se fundamental para aprimorar a tomada de decisões na educação.

3.1.7 Descoberta de conhecimento através da análise e mineração em dados do ENEM

O estudo (PORTO, 2019) focou na análise dos microdados do Enem 2017, buscando identificar padrões de desempenho, médias de notas e estatísticas por região do Brasil, além de explorar dados demográficos dos participantes, como idade, gênero, raça e escola de origem. Utilizando técnicas de mineração de dados e análise estatística, o objetivo foi compreender os fatores que influenciam o desempenho dos estudantes no exame. Os dados foram obtidos do portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP, permitindo análises detalhadas das notas, distribuição dos inscritos por região e comparações entre grupos demográficos.

O estudo seguiu uma metodologia detalhada, incluindo uma visão geral do ensino no Brasil e dos princípios do Enem, conceitos de banco de dados e técnicas de mineração de dados. Utilizou-se a biblioteca *Pandas*, biblioteca utilizada para manipulação e extração de dados, em *Python* para a análise, explorando diferentes perfis de participantes, como gênero, raça, renda e tipo de escola. A preparação dos dados foi crucial, dada a enorme quantidade de registros no banco de dados do Enem, exigindo a análise de amostras representativas para alcançar resultados viáveis.

Os resultados revelaram *insights* valiosos, identificando questões com altas taxas de acerto e erro, bem como comparativos de desempenho entre grupos demográficos. O estudo enfatizou a importância do Enem como avaliação nacional e sua relevância para o ensino médio, impactando diretamente o ingresso no ensino superior e no mercado de trabalho. Além disso, ofereceu um sistema de análise de microdados do Enem utilizando a biblioteca *Pandas*, apresentando uma abordagem eficaz para futuras análises educacionais.

3.1.8 Uso de *Business Intelligence* para avaliação de indicadores de desempenho na educação básica: um estudo de caso no estado do Acre

O estudo (WANDERLEY et al., 2021) se concentrou na criação de um sistema de *Business Intelligence* (BI) voltado para a educação, usando painéis interativos para exibir indicadores de desempenho na educação básica do estado do Acre. A quantidade massiva de dados na área educacional demanda uma ferramenta que traduza informações complexas em dados acessíveis para apoiar gestores na avaliação de políticas públicas e melhorias na educação. A integração de tecnologias de BI oferece uma visão sistêmica da educação, permitindo análises e *insights* para decisões mais informadas.

Os painéis desenvolvidos abrangem indicadores como distorção idade-série, esforço e formação docente, resultados em avaliações e evasão escolar permitindo aos usuários compreender claramente a situação educacional do Acre. A avaliação da ferramenta pelos usuários foi positiva, evidenciando sua eficácia e importância para os gestores educacionais.

O autor destacou a importância do BI na educação ao fornecer informações úteis para tomadas de decisão na gestão educacional, destacando também a necessidade de futuros estudos para ampliar a gama de indicadores, aprimorar o direcionamento de recursos e desenvolver modelos educacionais nacionais, utilizando técnicas avançadas de *machine learning*. *Machine learning*, ou aprendizado de máquina, é uma subárea da inteligência artificial que envolve a criação de algoritmos e modelos que permitem aos sistemas aprenderem e fazerem previsões ou tomar decisões com base em dados. Essas técnicas avançadas podem identificar padrões complexos nos dados educacionais e fornecer *insights* valiosos para melhorar o sistema educacional.

3.1.9 Comparativo entre os trabalhos correlatos

A Tabela 3.2 apresenta um comparativo dos artigos que compuseram esta seção, focando os aspectos das tecnologias utilizadas, base utilizada e resultados alcançados.

Artigo	Base de dados	Tecnologias Utilizadas	Resultados
Fatores Associados ao Desempenho de Alunos do 3º Ano do Ensino Médio do Distrito Federal: Uma Análise Multinível Aplicada a Dados do SAEB 2013	ANEB (Avaliação Nacional da Educação Básica), criado pela SAEB dos alunos do 3º ano do ensino médio.	Análise Multinível , regressão multinível, na preparação e análise dos dados foi utilizado a linguagem R	Com a escala de proficiência do SAEB que vai do nível 1 ao 10. Língua portuguesa o desempenho médio foi de 282,5, dentro da faixa do nível 3 da escala de proficiência. Enquanto para matemática, o desempenho médio foi de 292, também dentro da faixa do nível três da escala.
Fatores Associados ao Desempenho escolar em Matemática do SAEB 2011 para alunos do 5º ano no Distrito Federal	SAEB 2011 dos alunos do 5º ano do ensino fundamental no distrito federal. E o Censo Escolar 2011 com informações sobre a infraestrutura das escolas.	Análise Multinível, na preparação e análise dos dados foi utilizado a linguagem R	Fatores mais associados ao rendimento em Matemática de um aluno de 5º ano no Distrito Federal que prestou o SAEB em 2011 são o fato de ele nunca ter reprovado e ser um menino.
O que os dados do SAEB nos dizem sobre o desempenho dos estudantes em Matemática?	Dados utilizados do SAEB, disponíveis no Portal do INEP	Utilização de notas numéricas de proficiência, agrupadas em níveis para oferecer um panorama das habilidades matemáticas desenvolvidas e aquelas que demandam maior atenção	Aumento progressivo nas médias do 5º ano e oscilações constantes no 9º ano. No 3º ano do Ensino Médio, revela lacunas em habilidades essenciais.
Uma proposta de aplicação de <i>Business Intelligence</i> no sistema educacional brasileiro	O estudo utiliza dados do sistema educacional brasileiro disponibilizados pelo INEP e IDEB.	Uso do <i>Power BI</i> para análise de dados, incluindo etapas como seleção, ETL, <i>data warehouse</i> e descoberta de conhecimento.	O BI pode revelar informações valiosas nos dados educacionais do INEP, sendo útil para análises, decisões e implementações na educação.

Artigo	Base de dados	Tecnologias Utilizadas	Resultados
Desempenho Escolar E Escolha Entre Escolas Públicas E Privadas No Rio Grande Do Sul: Evidências Empíricas A Partir De Dados Do Saeb	SAEB	Medidas de estatística descritiva, elaboração de tabelas e gráficos.	Houve uma melhora de desempenho em Português e em Matemática no estado do Rio Grande do Sul, contudo não houve uma melhora significativa na posição observando-se a nível nacional, além disso notou-se melhor desempenho dos alunos das escolas privadas. Outros fatores como estrutura escolar, frequência na pré-escola, o maior nível social e a escolaridade superior da mãe estão relacionados com uma maior proficiência escolar.
A Construção De Um <i>Data Warehouse</i> Utilizando Os Indicadores Educacionais Do Inep	INEP e questionários realizados com escolas de Santa Maria - RS	Técnicas de modelagem de <i>data warehouse</i> , carregamento e visualização dos indicadores, uso de ferramenta de <i>BI</i> e análise da sua usabilidade	Em colaboração com a secretaria de educação local, o estudo validou a eficácia dos painéis, revelando resultados favoráveis para gestão educacional.
Descoberta De Conhecimento Através Da Análise E Mineração Em Dados Do Enem	ENEM 2017	Mineração de dados utilizando a biblioteca <i>Pandas</i> em <i>Python</i>	A análise dos microdados do Enem 2017 revelou padrões de desempenho significativos, destacando questões com altas taxas de acerto e erro, permitindo comparações entre grupos demográficos como gênero, raça, renda e tipo de escola. Além disso, forneceu um sistema que pode ser reaproveitado em trabalhos futuros.
Uso De <i>Business Intelligence</i> Para Avaliação De Indicadores De Desempenho Na Educação Básica: Um Estudo De Caso No Estado Do Acre	INEP, IBGE, SIOPE	<i>Python ETL</i> , <i>SQL Server</i> e <i>Power BI</i>	A avaliação da ferramenta pelos usuários foi positiva, evidenciando sua eficácia e importância para os gestores educacionais.

Tabela 3.2: Dados dos 8 artigos levantados

4

Desenvolvimento

Este capítulo tem como objetivo apresentar as etapas seguidas durante a realização do trabalho, cujo intuito é gerar informações que possam ser utilizadas na gestão escolar de escolas do Distrito Federal, utilizando dados do SAEB. Inicialmente, será exposta a metodologia da pesquisa aplicada, explicativa e experimental adotada neste estudo. Em seguida, será explicado como ocorreu a coleta e análise dos dados.

4.1 Metodologia

A metodologia adotada neste trabalho segue uma abordagem de pesquisa aplicada, explicativa e experimental, com o objetivo de desenvolver painéis que correlacionem as notas do SAEB, no período de 2013 a 2021, com as respostas dos questionários aplicados aos alunos do Distrito Federal.

O contexto em que essas abordagens metodológicas se aplicam ao TCC é o de proporcionar uma ferramenta prática e explicativa para os gestores educacionais do Distrito Federal, permitindo uma análise mais profunda e fundamentada das relações entre o desempenho dos alunos e os fatores investigados nos questionários. A pesquisa aplicada assegura que o produto final, os painéis, tenha uma utilidade prática direta; a pesquisa explicativa aprofunda o entendimento das correlações observadas; e a pesquisa experimental possibilita a exploração de diferentes abordagens para a visualização e interpretação dos dados.

Essas abordagens são amplamente defendidas na literatura acadêmica por sua capacidade de gerar conhecimento aplicável e útil para a prática educacional, conforme discutido por (CALEFFE; MOREIRA, 2006), que destacam a importância da pesquisa aplicada e explicativa na produção de soluções práticas e embasadas para problemas reais.

As ferramentas utilizadas para a realização dessas etapas foram *Python* e *Power BI*. A escolha do *Python* se deu pela sua robustez e versatilidade em manipulação de dados, além da ampla disponibilidade de bibliotecas específicas para transformação e limpeza de grandes volumes de informações, o que foi essencial para o tratamento dos dados do SAEB. O *Power BI* foi escolhido por sua capacidade de criar visualizações interativas e intuitivas, o que facilita a exploração e análise dos dados por parte dos gestores educacionais, transformando os *insights*

obtidos em informações acionáveis.

O *Python* foi empregado na parte de transformação e limpeza dos dados, por meio da criação de *scripts*, que são arquivos projetados para executar uma série de comandos de forma sequencial. Com isso, foi possível obter os dados referentes ao Distrito Federal, selecionar as perguntas relevantes e gerar arquivos com essas informações. Após essa etapa, utilizou-se o *Power BI* para a construção dos painéis.

4.2 Coleta e Análise de Dados

A coleta da base de dados do Sistema de Avaliação da Educação Básica (SAEB) teve início em outubro de 2023. Inicialmente, procurou-se consolidar um conjunto de dados abrangentes sobre os alunos do Distrito Federal. Para isso, foram utilizados os microdados do SAEB de 2013 a 2021, que contêm informações sobre alunos, escolas, professores, diretores e secretários municipais. Esses dados podem ser acessados pelo portal do INEP conforme visto na figura 4.1, disponível em (Microdados SAEB, 2024).

Figura 4.1: Como Obter os Microdados SAEB

A imagem mostra a interface do portal gov.br. No topo, há o logotipo do gov.br e o nome do Ministério da Educação. Abaixo, há links para "Órgãos do Governo", "Acesso à Informação", "Legislação" e "Acessibilidade". Um botão azul "Entrar com o gov.br" está no canto superior direito. O endereço do site é "Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep". Uma barra de busca contém o texto "O que você procura?". O caminho de navegação é "Acesso à Informação > Dados Abertos > Microdados > Saeb". O título principal é "Saeb" e o subtítulo é "Microdados do Sistema de Avaliação da Educação Básica". Abaixo, há informações de publicação: "Publicado em 16/09/2022 10h28" e "Atualizado em 08/12/2023 15h20". Há ícones para compartilhar em redes sociais. O conteúdo principal mostra uma lista de microdados por ano:

- 2021
 - Microdados Saeb 2021 - Ensino fundamental e médio (Atualizado em 05/05/2023)
 - Microdados Saeb 2021 - Educação Infantil
- 2019
 - Microdados do Saeb 2019
- 2017
 - Microdados do Saeb 2017

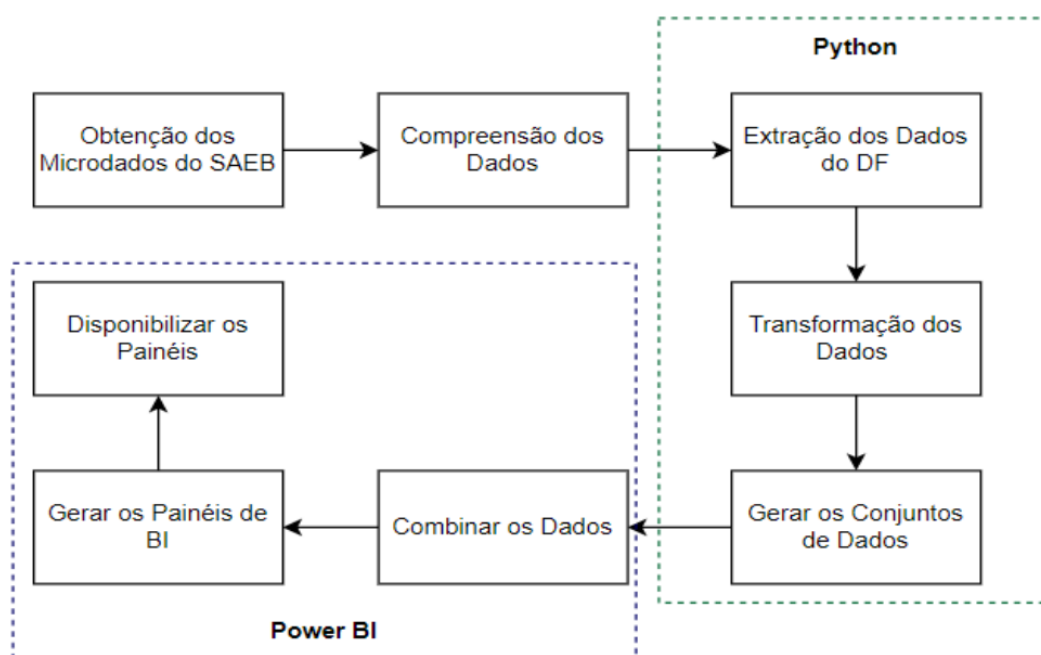
Fonte: (Microdados SAEB, 2024)

O tratamento e limpeza dos dados foi uma etapa crucial para garantir a integridade e a consistência das análises realizadas. Inicialmente, foram removidas variáveis consideradas irrelevantes ou redundantes para os objetivos do estudo. Em seguida, foi realizada a normalização dos dados, unificando escalas de notas e padronizando as respostas dos questionários para permitir uma análise comparativa ao longo dos anos.

Uma das principais dificuldades enfrentadas foi a necessidade de harmonizar variáveis que mudaram entre as edições do SAEB. Por exemplo, algumas perguntas dos questionários foram reformuladas ou substituídas em determinados anos, exigindo ajustes específicos para manter a comparabilidade dos dados. Para isso, aplicamos técnicas de recategorização e normalização. A recategorização envolve a reclassificação de categorias de respostas para que variáveis semelhantes sejam agrupadas de maneira consistente ao longo do tempo, mesmo quando originalmente apresentavam diferenças. A normalização dos dados, por sua vez, ajusta os valores das variáveis para uma escala comum, facilitando a comparação direta entre diferentes anos e questionários.

Além disso, valores ausentes ou inconsistentes foram tratados de acordo com métodos estabelecidos na literatura, como a imputação de valores médios ou a exclusão de registros incompletos, dependendo do caso. Todo o processo foi conduzido utilizando *scripts* em *Python*, o que permitiu um controle rigoroso e a replicabilidade das operações realizadas. O processo de coleta e análise dos microdados do SAEB foi estruturado em duas etapas principais, conforme ilustra o diagrama 4.2. Primeiramente, a obtenção dos microdados do SAEB e a compreensão dos dados disponíveis foram essenciais para guiar a extração específica dos dados relacionados ao DF, que foi realizada utilizando *scripts* em *Python*. Após essa extração, foi feita a transformação dos dados para garantir a consistência e padronização, gerando os conjuntos de dados necessários para a análise. Em seguida, esses conjuntos de dados foram combinados com outros dados relevantes e carregados no *Power BI*, onde foram criados e disponibilizados os painéis de BI. Esses painéis visam fornecer uma visão abrangente dos resultados educacionais, facilitando a tomada de decisões informada por parte dos gestores escolares.

Figura 4.2: Fluxo de Processamento e Visualização dos Dados do SAEB



Para a análise dos dados do SAEB referente aos alunos do 5º e 9º ano do Ensino Fundamental e do 3º ano do Ensino Médio, foram aplicados diversos filtros para selecionar as informações relevantes. Inicialmente, utilizou-se o ID correspondente ao Distrito Federal (ID_UF = 53) e foram filtrados apenas os registros onde a prova e o questionário foram devidamente preenchidos (IN_PREENCHIMENTO_PROVA = 1 e IN_PREENCHIMENTO_QUESTIONARIO = 1). Além disso, foram removidos registros com valores de proficiência inválidos (PROFICIENCIA_LP_SAEB <> "nan").

Para padronizar os dados, o identificador de prova foi renomeado nos anos de 2019 e 2021 de ID_SAEB para ID_PROVA_BRASIL, alinhando-o com os outros anos de análise. Apenas as colunas de interesse, como proficiência em Matemática e Língua Portuguesa e outras variáveis essenciais para a correlação dos dados, foram mantidas na análise. Essas colunas incluíam, por exemplo, ID_PROVA_BRASIL, ID_ESCOLA, e PROFICIENCIA_LP_SAEB.

Ao final, foram selecionadas as seguintes perguntas dos questionários aplicados aos alunos:

- A escola é pública?
- Qual a Localização da Escola?
- Qual é a sua cor ou raça?
- Na sua casa tem televisão?
- Na sua casa tem geladeira?
- Na sua casa tem máquina de lavar roupa?
- Na sua casa tem carro?
- Na sua casa tem computador?
- Na sua casa tem banheiro?
- Na sua casa tem quartos para dormir?
- Normalmente, quem mora na sua casa? - Mãe (mães ou madrasta).
- Qual é a maior escolaridade da sua mãe (ou mulher responsável por você)?
- Normalmente, quem mora na sua casa? - Pai (pais ou padrasto).
- Qual é a maior escolaridade de seu pai (ou homem responsável por você)?
- Seus pais ou responsáveis vão à reunião de pais?
- Fora da escola em dias de aula, quanto tempo você usa para: - Lazer (TV, internet, brincar, música etc.).

- Fora da escola em dias de aula, quanto tempo você usa para: - Trabalhar em casa (lavar louça, limpar quintal, cuidar dos irmãos).
- Você já foi reprovado(a)?
- Alguma vez você abandonou a escola deixando de frequentá-la até o final do ano escolar?
- Seus pais ou responsáveis incentivam você a estudar?
- Seus pais ou responsáveis incentivam você a fazer o dever de casa e/ou os trabalhos da escola?
- Seus pais ou responsáveis incentivam você a ir a escola e/ou não faltar às aulas?
- Seus pais ou responsáveis conversam com você sobre o que acontece na escola?
- Atualmente você trabalha fora de casa (recebendo ou não um salário)?
- Na sua casa tem: - Freezer (independente ou segunda porta da geladeira).

Para mais informações detalhadas sobre as etapas de obtenção, tratamento e limpeza dos dados, bem como para acessar os *scripts* utilizados no processo, os interessados podem visitar o (Repositório SAEB, 2024) disponível no *GitHub*. O repositório contém a documentação completa, além dos códigos em *Python* que foram empregados para a análise dos microdados do SAEB.

5

Resultados

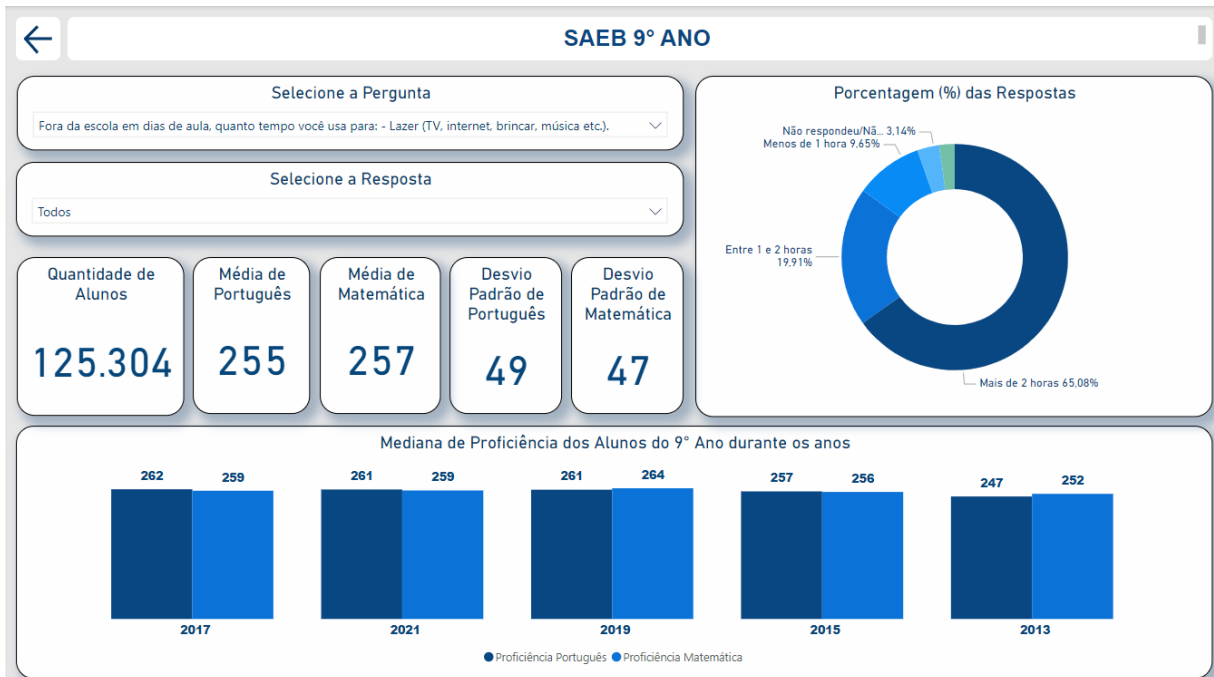
O capítulo de resultados apresenta as análises realizadas a partir dos dados do SAEB e das respostas dos questionários aplicados aos alunos do Distrito Federal e também as limitações da pesquisa. Para facilitar a visualização e interpretação desses dados, foi desenvolvido um painel interativo no *Power BI*. Este painel permite aos gestores educacionais explorar a relação entre as respostas obtidas nos questionários e as proficiências dos alunos no SAEB, além de oferecer uma visão detalhada do desempenho das escolas de acordo com os grupos socioeconômicos. Além disso, é possível visualizar a quantidade de alunos e escolas do DF, fornecendo uma visão abrangente do panorama educacional da região conforme visto nas figuras 5.1, 5.2 e 5.3. O uso do *Power BI* nesse processo foi fundamental para a criação de uma ferramenta prática e intuitiva, que auxilia na tomada de decisões informadas e na identificação de áreas que requerem atenção específica.

Figura 5.1: Exemplo de uso do Painel do 5º ano



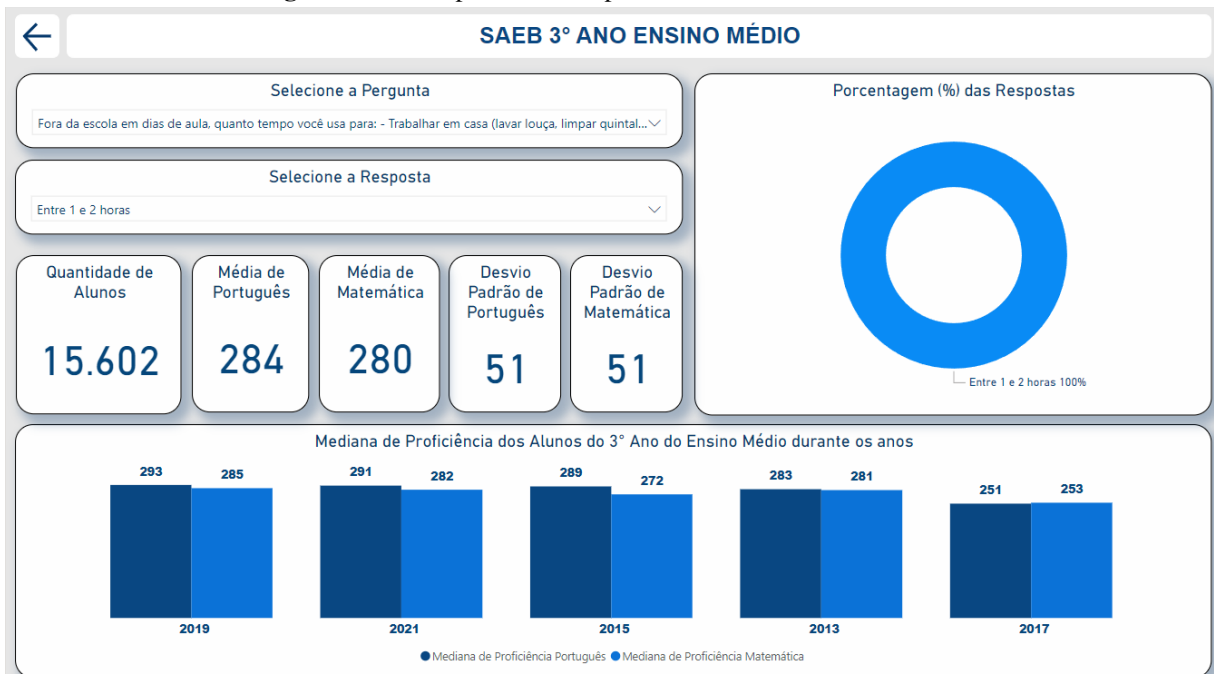
Fonte: Autor

Figura 5.2: Exemplo de uso do painel do 9º ano



Fonte: Autor

Figura 5.3: Exemplo de uso do painel do 3º ano do ensino médio

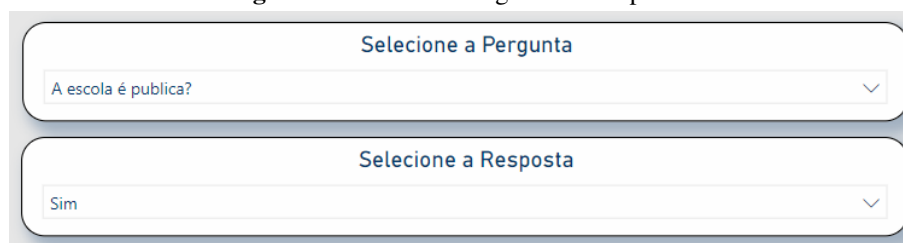


Fonte: Autor

Nos painéis referentes ao 5º ano, 9º ano e 3º ano do Ensino Médio, a pergunta a ser investigada pode ser selecionada na parte superior esquerda. Em seguida, é possível escolher uma das respostas disponíveis. O painel exibirá a quantidade de alunos que deram essa resposta, as médias de proficiência em Português e Matemática, e os respectivos desvios padrão. À

direita, encontra-se a porcentagem de respostas para cada opção de pergunta, e na parte inferior, a mediana das proficiências é apresentada de acordo com os anos, permitindo uma análise comparativa detalhada. Conforme visto nos exemplos a seguir (Figura 5.4 e 5.5),

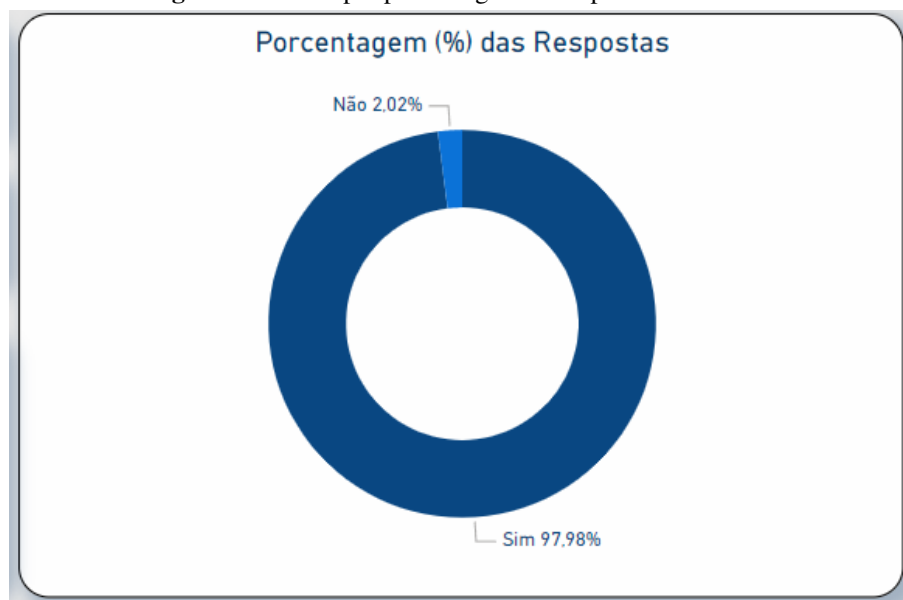
Figura 5.4: Filtro de Perguntas e Respostas



The image shows a user interface for filtering data. It consists of two stacked dropdown menus. The top menu is titled "Selecione a Pergunta" and currently displays "A escola é publica?". The bottom menu is titled "Selecione a Resposta" and currently displays "Sim". Both menus have a downward-pointing arrow on the right side.

Fonte: Autor

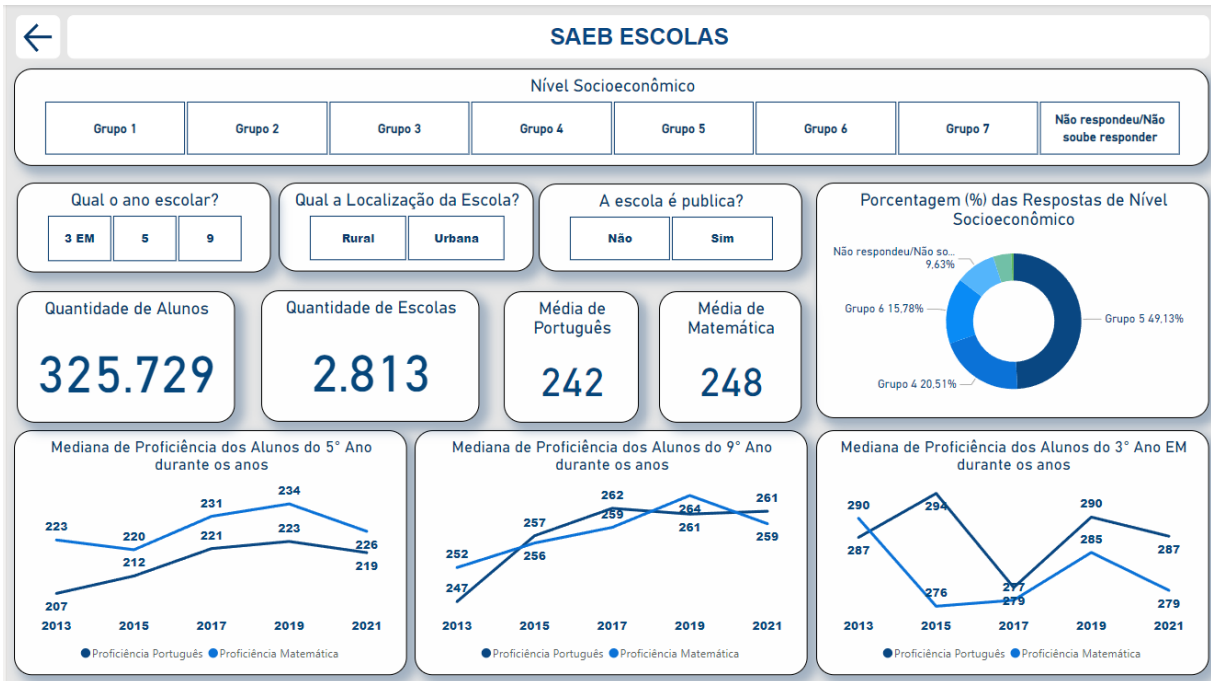
Figura 5.5: Exemplo porcentagem da resposta selecionada



Fonte: Autor

No painel de escolas, é apresentado na parte superior um filtro para grupos socioeconômicos conforme a classificação realizada pelo SAEB, onde o Grupo 1 representa o menor nível socioeconômico e o Grupo 7 o maior. Esse indicador não se baseia apenas na renda familiar, mas também inclui outros parâmetros importantes, como o nível educacional dos pais e a posse de bens e serviços no domicílio do estudante. Esses fatores são utilizados para compor o Índice Socioeconômico (Inse) do Saeb.

Figura 5.6: Painel de escolas



Fonte: Autor

Figura 5.7: Exemplo de uso do painel de escolas



Fonte: Autor

Figura 5.8: Filtro Nível socioeconômico

Nível Socioeconômico							
Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Não respondeu/Não soube responder

Fonte: Autor

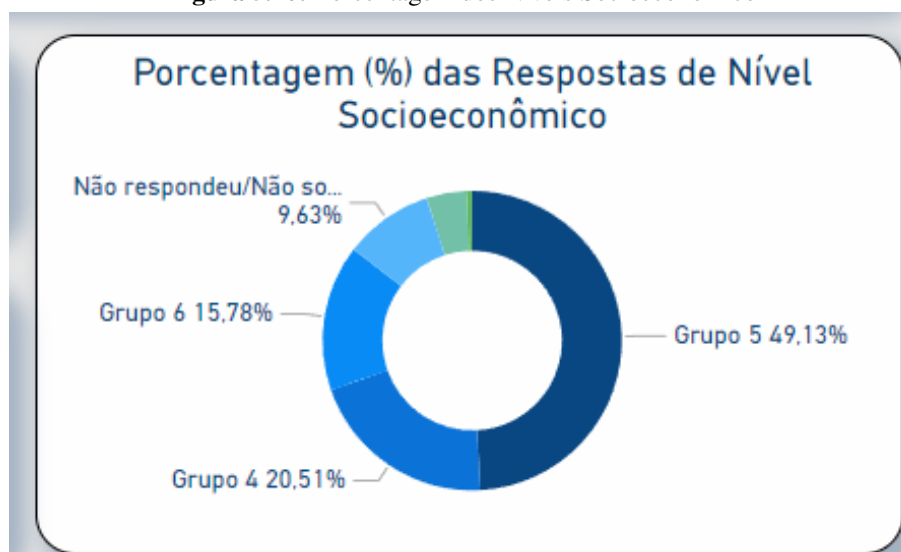
Após a seleção do grupo, é possível escolher o ano escolar, a localização da escola (rural ou urbana) e o tipo de gestão (pública ou privada).

Figura 5.9: Filtro de Ano, Localização e Tipo de Escola

<p>Qual o ano escolar?</p> <table border="1"> <tr> <td>3 EM</td> <td>5</td> <td>9</td> </tr> </table>	3 EM	5	9	<p>Qual a Localização da Escola?</p> <table border="1"> <tr> <td>Rural</td> <td>Urbana</td> </tr> </table>	Rural	Urbana	<p>A escola é pública?</p> <table border="1"> <tr> <td>Não</td> <td>Sim</td> </tr> </table>	Não	Sim
3 EM	5	9							
Rural	Urbana								
Não	Sim								

Fonte: Autor

Além disso, o painel exibe a porcentagem de cada grupo, a quantidade de alunos, o número de escolas e as proficiências correspondentes ao grupo selecionado.

Figura 5.10: Porcentagem dos Níveis Socioeconômico

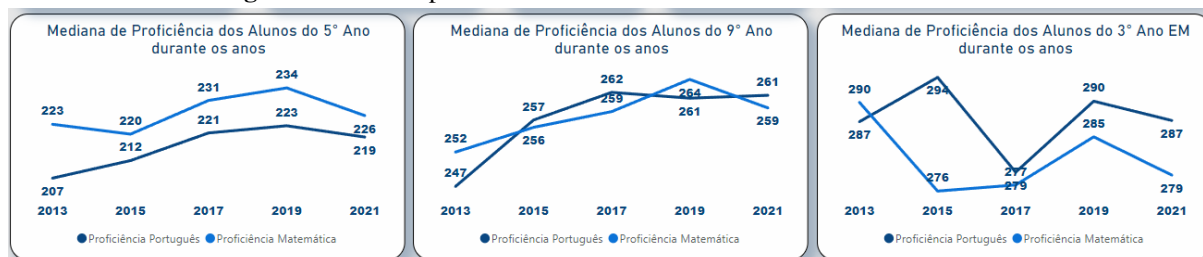
Fonte: Autor

Figura 5.11: Exemplos de informações contidas no Painel de Escolas

Fonte: Autor

Na parte inferior, é possível visualizar a variação da mediana ao longo dos anos entre 2013 e 2021.

Figura 5.12: Exemplo de Mediana de Acordo com o Ano de 2013 a 2021



Fonte: Autor

Os painéis de BI desenvolvidos oferecem aos gestores escolares uma ferramenta robusta para a tomada de decisões informadas. Através da análise detalhada de dados socioeconômicos, desempenho dos alunos e variações ao longo do tempo, os gestores podem identificar padrões, acompanhar a evolução das proficiências e alocar recursos de forma mais eficaz. Além disso, a visualização clara das informações permite um entendimento mais profundo das realidades das escolas, facilitando a implementação de estratégias que visem à melhoria da qualidade educacional e à redução das desigualdades no sistema de ensino.

5.1 Limitações da Pesquisa

Uma das principais limitações desta pesquisa está relacionada ao uso de códigos fictícios aplicados aos microdados do SAEB para atender às exigências da Lei Geral de Proteção de Dados (LGPD). Devido a essa anonimização, não foi possível identificar as regiões administrativas do Distrito Federal, o que limitou a análise geográfica dos dados. Além disso, nos anos de 2019 e 2021, foi identificado que 5.065 alunos não estavam vinculados a nenhuma escola na tabela de escolas, o que impediu a inclusão desses dados em certas análises. Para manter a integridade do painel de escolas, foi aplicado um filtro no nível socioeconômico para excluir esses alunos. Outro desafio enfrentado foi a necessidade de descartar algumas perguntas dos questionários, que sofreram modificações ao longo dos anos, dificultando a comparação histórica. Por fim, uma limitação técnica do *Power BI* foi a dificuldade em selecionar simultaneamente múltiplas respostas ou perguntas, o que restringiu algumas análises mais complexas.

6

Conclusões

Este trabalho demonstrou a importância do Sistema de Avaliação da Educação Básica (SAEB) como uma ferramenta fundamental para avaliar a qualidade do ensino no Brasil, em especial no Distrito Federal. Através da análise dos dados do SAEB, cobrindo o período de 2013 a 2021, e da criação de painéis de *Business Intelligence* (BI), foi possível fornecer aos gestores educacionais uma visão mais detalhada e fundamentada sobre a relação entre o desempenho dos alunos e os fatores socioeconômicos investigados nos questionários aplicados.

A pesquisa seguiu uma abordagem aplicada, explicativa e experimental, resultando em um produto final que vai além da análise estatística, propondo uma ferramenta prática para o apoio à gestão escolar. O desenvolvimento dos painéis de BI, acessíveis através de um Sistema Web, permite não só a exploração das correlações entre as proficiências dos alunos e suas condições socioeconômicas, mas também uma visão global sobre o desempenho das escolas, facilitando a identificação de áreas críticas e auxiliando na tomada de decisões estratégicas.

As ferramentas utilizadas, *Python* e *Power BI*, foram escolhidas por sua robustez e versatilidade, desempenhando papéis cruciais no tratamento dos microdados do SAEB e na criação de visualizações interativas. Essas visualizações facilitaram o acesso e a compreensão das análises pelos gestores educacionais. A combinação das técnicas de tratamento de dados com as funcionalidades do Power BI assegurou que as informações fossem apresentadas de maneira precisa, relevante e fácil de utilizar na tomada de decisões.

No entanto, a pesquisa enfrentou algumas limitações, como a impossibilidade de identificar as regiões administrativas do DF devido à anonimização dos dados para atender à LGPD, e a necessidade de descartar certas perguntas dos questionários ao longo dos anos para assegurar a comparabilidade histórica. Mesmo assim, o painel desenvolvido se mostrou uma ferramenta valiosa, proporcionando uma visão abrangente e detalhada da educação no Distrito Federal.

Para trabalhos futuros, recomenda-se o desenvolvimento de funcionalidades avançadas nos painéis de BI, como a implementação de algoritmos de *machine learning* para a identificação automática de padrões e tendências. Além disso, explorar a integração desses painéis com outras bases de dados educacionais e utilizar outros questionários, poderia enriquecer ainda mais as análises. Outro ponto relevante seria a investigação de metodologias que permitam superar as limitações impostas pela anonimização dos dados, buscando alternativas para incluir de forma

segura as regiões administrativas do Distrito Federal, o que possibilitaria uma análise geoespacial mais detalhada.

Finalmente, este estudo contribui de maneira significativa para o campo da gestão educacional, oferecendo uma ferramenta que não apenas facilita a análise dos dados do SAEB, mas também pode ser adaptada e utilizada em outras regiões ou contextos educacionais. A continuidade deste trabalho pode explorar novas abordagens de visualização e expandir as análises para incluir variáveis adicionais, aprofundando ainda mais o entendimento sobre os fatores que influenciam a qualidade da educação no Brasil.

Referências

- BATISTA, G. M. **Fatores associados ao desempenho de alunos do 3º ano do ensino médio do Distrito Federal** : uma análise multinível aplicada a dados do saeb 2013. [S.l.]: Universidade de Brasília, 2016. Disponível em <<https://bdm.unb.br/handle/10483/15171>>. Acessado em Novembro 2023.
- CALEFFE, L. G.; MOREIRA, H. Metodologia da pesquisa para o professor pesquisador. **ILJ**, [S.l.], v.17, p.1, 2006.
- Cetax. **Data Warehouse / Armazém de Dados**: o que é, conceito e definição. Acessado em Outubro 2023.
- Corporate Finance Institute. **Data Warehousing**. Acessado em Junho 2024, <https://corporatefinanceinstitute.com/resources/business-intelligence/data-warehousing/>.
- FITRIANA, R.; ERIYATNO, T. D.; DJATNA, T. Progress in Business Intelligence System research: a literature review. **International Journal of Basic & Applied Sciences IJBAS-IJENS**, [S.l.], v.11, n.03, p.118503–6464, 2011.
- GILAD, B.; GILAD, T. A systems approach to business intelligence. **Business Horizons**, [S.l.], v.28, n.5, p.65–70, 1985.
- HABUL, A.; PILAV-VELIC, A. Business intelligence and customer relationship management. In: ITI 2010, 32ND INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY INTERFACES. **Proceedings...** [S.l.: s.n.], 2010. p.169–174.
- ILHA, L. B. et al. A construção de um data warehouse utilizando os indicadores educacionais do INEP. , [S.l.], 2021. Acessado em Dezembro 2023.
- INEP. **Nota Explicativa**: resultados prova brasil 2013. Acessado em Novembro 2023, https://www.seed.se.gov.br/provabrasil/arquivos/ProvaBrasil/Resultado_Prova_Brasil_2013.pdf.
- JUNIOR, C. G. M. **UMA PROPOSTA DE APLICAÇÃO DE BUSINESS INTELLIGENCE NO SISTEMA EDUCACIONAL BRASILEIRO**. [S.l.]: Universidade Federal do Amazonas, 2021. Disponível em <<https://tede.ufam.edu.br/bitstream/tede/8655/9/Disserta>>
- LIMA AMBROSIO Érica. **Fatores associados ao desempenho escolar em matemática do SAEB 2011 para alunos do 5º ano no Distrito Federal** : um estudo multinível. [S.l.]: Universidade de Brasília, 2014. Disponível em <<https://bdm.unb.br/handle/10483/8716>>. Acessado em Novembro 2023.
- MENEZES, E. T. de. **Dicionário Interativo da Educação Brasileira - EducaBrasil**. [S.l.]: EDUCABRASIL, 2001. Disponível em <<https://www.educabrasil.com.br/sistema-educacional-brasileiro>>. Acessado em Novembro 2023.
- Microdados SAEB. **Microdados SAEB**. Acessado em Junho 2024, <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/saeb>.

- Microsoft. **O que é o Power BI?** Acessado em Outubro 2023, <https://powerbi.microsoft.com/pt-br/what-is-power-bi/>.
- Microsoft. **Excel.** Acessado em Outubro 2023, <https://www.microsoft.com/pt-br/microsoft-365/excel>.
- PEREIRA, J. B.; TOLENTINO-NETO, L. C. B. d. O que os dados do SAEB nos dizem sobre o desempenho dos estudantes em Matemática? **Educação Matemática Pesquisa**, [S.l.], v.17, n.2, p.309–333, 2015.
- PORTO, K. G. S. Descoberta de conhecimento através da análise e mineração em dados do Enem. , [S.l.], 2019. Acessado em Dezembro 2023.
- PRESTES, G. S. P. **Desempenho escolar e escolha entre escolas públicas e privadas no Rio Grande do Sul: evidências empíricas a partir de dados do saeb.** 2016. Dissertação (Mestrado em Ciência da Computação) — . Acessado em Dezembro 2023.
- Python Software Foundation. **Getting Started.** Acessado em Outubro 2023, <https://www.python.org>.
- Repositório SAEB. **Repositório.** Acessado em Junho 2024, <https://github.com/OliveiraUP/TCC>.
- SAEB. **Questionários e testes SAEB.** Acessado em Junho 2024, <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb/testes-e-questionarios>.
- SAEB. **Sistema de Avaliação da Educação Básica.** Acessado em Novembro 2023, <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>.
- TURBAN, E. et al. **Business Intelligence: um enfoque gerencial para a inteligência do negócio.** [S.l.]: Bookman Editora, 2009.
- WANDERLEY, P. F. et al. Uso de business intelligence para avaliação de indicadores de desempenho na educação básica: um estudo de caso no estado do acre. , [S.l.], 2021. Acessado em Dezembro 2023.