



Construção do Data Mart do IDEB

Discentes: Ana Paula Pereira e Silva | ana.ppsilva2@ufrpe.br

Lucas Matheus Almeida Pereira | lucas.matheusalmeida@ufrpe.br

Nicolas de Moraes Borges Dória | nicolas.borges@ufrpe.br

Fonte dos Dados: <[Base dos Dados – IDEB \(INEP/MEC\)](#)>

Aplicação OLAP: <https://app.powerbi.com/links/_rYFMWT4or?ctid=a6f74efa-3674-449a-acae-coa23>

Github: [lucasmatheusalmeida/sad_ideb](https://github.com/lucasmatheusalmeida/sad_ideb)

ETAPA 1. PLANEJAMENTO

1. Contextualização

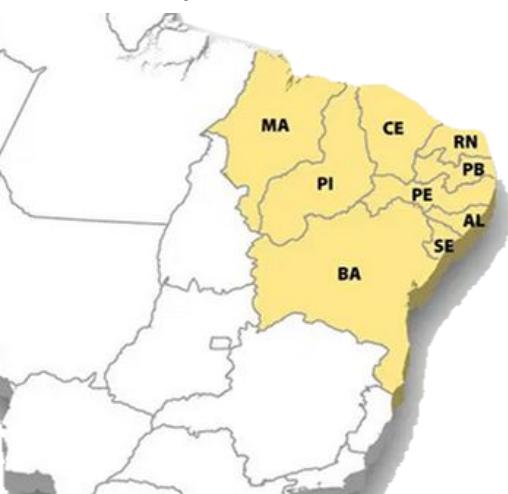
O IDEB é o principal indicador da qualidade da educação básica no Brasil, combinando o fluxo escolar — como taxas de aprovação — com o desempenho dos estudantes nas avaliações do SAEB (DEMO; SILVA; MINAYO, 2021).

A **Região Nordeste** historicamente apresenta os menores índices educacionais do país, mas vem mostrando evolução significativa. Compreender essa trajetória entre 2005 e 2021 é essencial para identificar políticas que funcionam e onde precisamos avançar.

Mesmo com o progresso, a região ainda enfrenta desafios importantes, como maiores taxas de evasão, desigualdades entre as redes pública e privada e diferenças de desempenho entre os anos iniciais, anos finais e o ensino médio. Estudos recentes destacam que iniciativas como escolas de tempo integral têm potencial para reduzir parte dessas desigualdades (DEMO; SILVA; MINAYO, 2021).

2. Escopo/objetivo do Data Mart

O objetivo do Data Mart a ser desenvolvido é fornecer uma análise detalhada da **evolução do IDEB na Região Nordeste** em diferentes redes de ensino e etapas escolares ao longo do tempo. O Data Mart permitirá extrair insights valiosos sobre as tendências de qualidade educacional, identificar fatores associados ao desempenho e avaliar a eficácia de políticas públicas educacionais.



Os dados a serem analisados serão as **redes de ensino** (pública, privada, estadual e municipal), as **etapas escolares** (anos iniciais, anos finais, ensino médio) e os **indicadores educacionais** (IDEB, taxa de aprovação, notas SAEB). O período de análise são os anos 2005, 2007, 2009, 2011, 2013, 2015, 2017, 2019 e 2021.

3. Arquitetura Tecnológica

Fonte de Dados:



Parte de ETL:



Aplicação OLAP:



4. Processo

- Planejamento
- Levantamento das necessidades
- Modelagem dimensional
- Projeto físico dos BD's
- Projeto ETC
- Desenvolvimento e Aplicações OLAP



5. Abordagem

Para o desenvolvimento deste Data Mart foi utilizada a abordagem **Bottom-up**, pois ela é uma abordagem evolutiva, com disponibilização imediata dos produtos. Como se trata de um Data Mart específico do IDEB, que foca em uma área restrita (Região Nordeste e seus indicadores), é mais adequado utilizar a abordagem Bottom-up, pois parte das necessidades locais para construir algo maior, se necessário.

Na modelagem, foi adotado o **Star Schema**, que é amplamente utilizado na construção de Data Marts por sua simplicidade e eficiência. Esse modelo organiza os dados em tabelas de dimensão e fato, facilitando o entendimento, a consulta e a análise. Diferentemente de arquiteturas mais centralizadas, o Star Schema promove acesso direto e rápido às informações essenciais, o que o torna apropriado para análises educacionais voltadas ao IDEB.

6. Usuários

- Gestores de instituições de ensino
- Órgãos públicos e secretarias de educação
- Pesquisadores e analistas de educação
- Áreas de Business Intelligence e Analytics
- Estudantes e sociedade em geral
- Empresas privadas do setor educacional
- Órgãos de controle e fiscalização



ETAPA 2.

LEVANTAMENTO DAS NECESSIDADES

7. Consultas de Apoio à Decisão

- Evolução do IDEB ao longo dos anos na Região Nordeste
- Comparação do IDEB entre redes de ensino (pública, privada, estadual) por ano
- Comparação do IDEB entre etapas de ensino (anos iniciais, finais, ensino médio) por ano
- Distribuição da taxa de aprovação e das notas SAEB por rede e etapa de ensino ao longo do tempo



8. Indicadores da análise

- Evolução Histórica no Nordeste
- Comparativo Nordeste vs Outras Regiões
- IDEB médio por rede de ensino
- IDEB médio por etapa de ensino
- Média do IDEB por Ano com a Rede de Ensino
- Comparativo da Média do IDEB no Nordeste Vs Brasil
- Comparativo das Notas do Nordeste vs Brasil
- Taxa de aprovação (%) por rede e etapa de ensino
- Relação entre IDEB e taxa de aprovação
- Evolução percentual do IDEB ao longo dos anos

ETAPA 3. MODELAGEM

9. Modelo Relacional

Tempo
id
ano

Rede Ensino
id
rede_ensino
rede

Etapa Ensino
id
etapa_ensino
etapa

Região
id
regiao

10. Modelo Dimensional

- Atributos e Hierarquia das Dimensões
 - dim_tempo (ano)
 - dim_rede_ensino (rede)
 - dim_etapa_ensino (etapa)
 - dim_regiao(regiao)

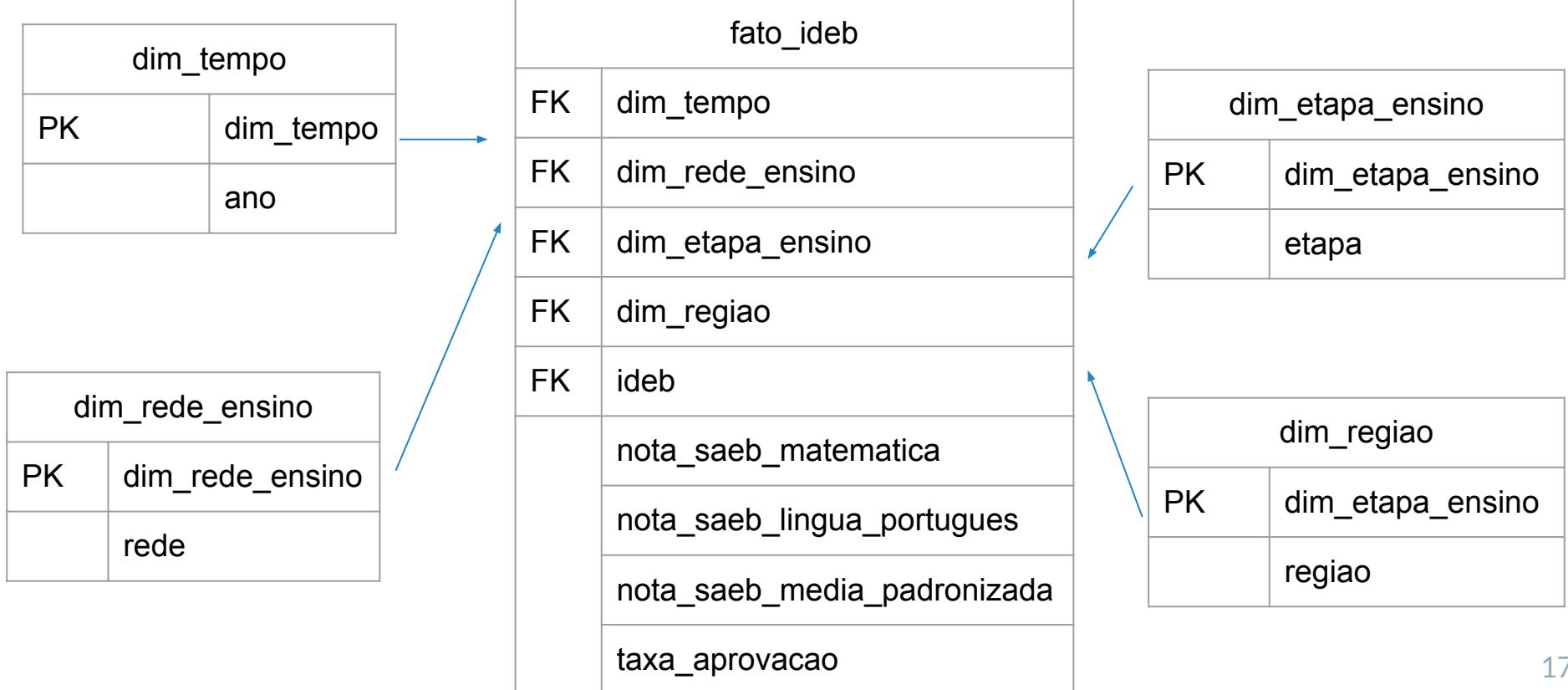
10. Modelo Dimensional

- Área de Negócios
 - Educação
- Processo
 - Avaliar evolução do IDEB no Nordeste
- Granularidade
 - Rede × Etapa × Regiao × Ano

10. Modelo Dimensional

- Métricas da Fato
- fato_ideb
 - ideb (Aditiva)
 - taxa_aprovacao (Aditiva)
 - nota_saeb_media_padronizada (Aditiva)
 - nota_saeb_matematica (Aditiva)
 - nota_saeb_lingua_portuguesa (Aditiva)

10. Modelo Dimensional



10. Modelo Dimensional

dim_rede_ensino	
PK	rede_ensino
1	publica
2	privada
3	estadual
4	total

dim_tempo	
PK	tempo
1	2005
2	2007
3	2009
4	2011
5	2013

dim_etapa_ensino	
PK	etapa_ensino
1	fundamental1
2	fundamental2
3	medio

dim_regiao	
PK	regiao
1	Nordeste
2	Sudeste
3	Sul

10. Modelo Dimensional

id_rede	id_tempo	id_etapa	id_regiao	id_ideb	taxa_aprovacao	nota_matematica	nota_portuguesa	nota_padronizada
1	1	1	1	3	73,7	260	248	40645
2	2	2	2	3.3	7,5	232	226	43243
3	3	3	3	3.9	6,0	181	172	45518
4	4	3	2	3.1	97,4	228	222	41828
4	5	3	1	3.4	80	174	165	43024
2	3	5	2	3.5	81,6	272	261	47005
3	3	1	3	3.8	56.2	247	234	44352
2	2	3	2	4.2	84	193	175	48566
4	5	3	1	4.5	88	261	365	42353
1	5	1	2	4.9	77,6	300	200	40000

11. Modelo Dimensional do Data Mart

- Estimativa de Espaço

O período considerado é de 9 anos (2005 a 2021), com 500 registros por dia.

$9 \times 365 = 3.285$ dias, resultando em:

$3285 \times 500 = 1.642.500$ registros.

Cada registro possui 4 chaves estrangeiras (pk_tempo, pk_rede_ensino, pk_etapa_ensino, pk_regiao), com 4 bytes cada, e 4 medidas numéricas (ideb, nota_saeb_matemática, nota_saeb_portuguesa e nota_saeb_média_padronizada), também com 4 bytes cada.

Assim, temos:

$$1.642.500 \times ((3 \times 4) + (4 \times 4)) = 1.642.500 \times 28 = 52.560.000 \text{ bytes} = 52.560 \text{ KB.}$$

11. Modelo Dimensional do Data Mart

- Estimativa de Espaço
 - ▷ Adicionado 20% referentes às dimensões e índices:
20% de 52.560 KB = 10.512 KB.
 - ▷ Total estimado:
 $52.560 + 10.512 = 63.072 \text{ KB}$ (aproximadamente 52,6 MB).

ETAPA 4. PROJETO FÍSICO DO DB

12. Modelo Relacional do Data Mart (Físico)

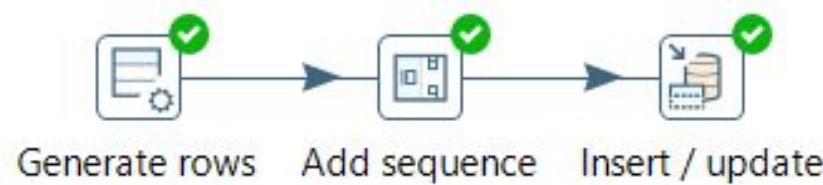
```
1  CREATE TABLE dim_etapa_ensino (
2      id SERIAL PRIMARY KEY,
3      etapa VARCHAR(100),
4      anos_escolares VARCHAR(50)
5  );
6
7  CREATE TABLE dim_regiao (
8      id SERIAL PRIMARY KEY,
9      regiao VARCHAR(50) NOT NULL UNIQUE
10 );
11
12 CREATE TABLE dim_tempo (
13     id SERIAL PRIMARY KEY,
14     ano INTEGER
15 );
16
17 CREATE TABLE dim_rede_ensino (
18     id SERIAL PRIMARY KEY,
19     rede_ensino VARCHAR(100) NOT NULL
20 );
```



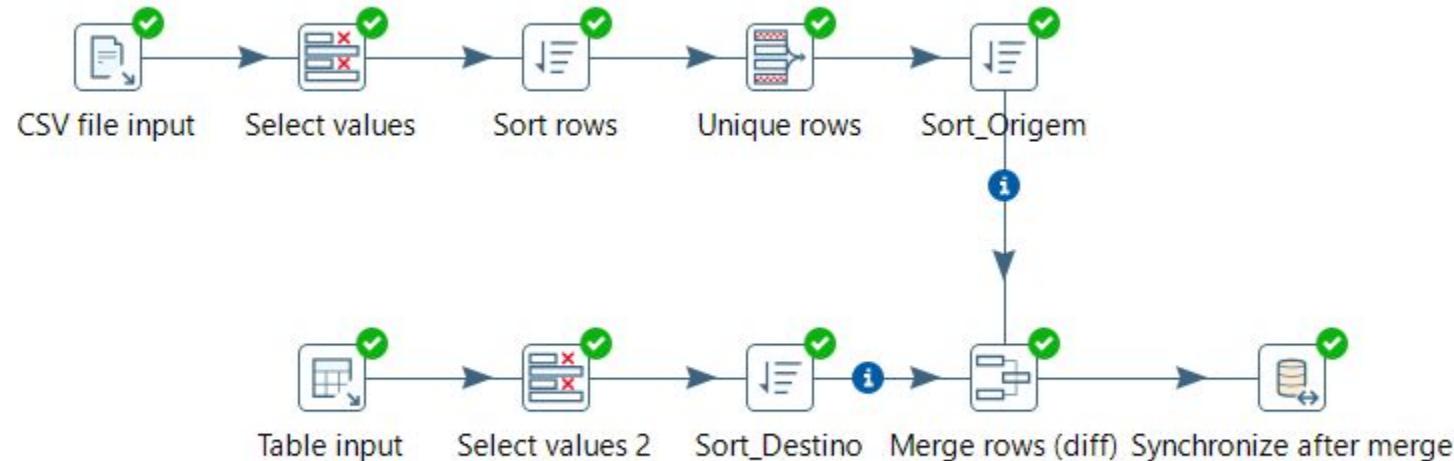
```
22  CREATE TABLE fato_ideb (
23      id_tempo INTEGER REFERENCES dim_tempo(id),
24      id_rede_ensino INTEGER REFERENCES dim_rede_ensino(id),
25      id_etapa INTEGER REFERENCES dim_etapa_ensino(id),
26      id_regiao INTEGER REFERENCES dim_regiao(id),
27      taxa_aprovacao NUMERIC(5,1),
28      nota_saeb_matematica NUMERIC(6,2),
29      nota_saeb_lingua_portuguesa NUMERIC(6,2),
30      nota_saeb_media_padronizada NUMERIC(6,2),
31      ideb NUMERIC(3,1),
32      PRIMARY KEY (id_tempo, id_rede_ensino, id_etapa, id_regiao)
33 );
```

ETAPA 5. EXTRAÇÃO, TRANSFORMAÇÃO E CARGA

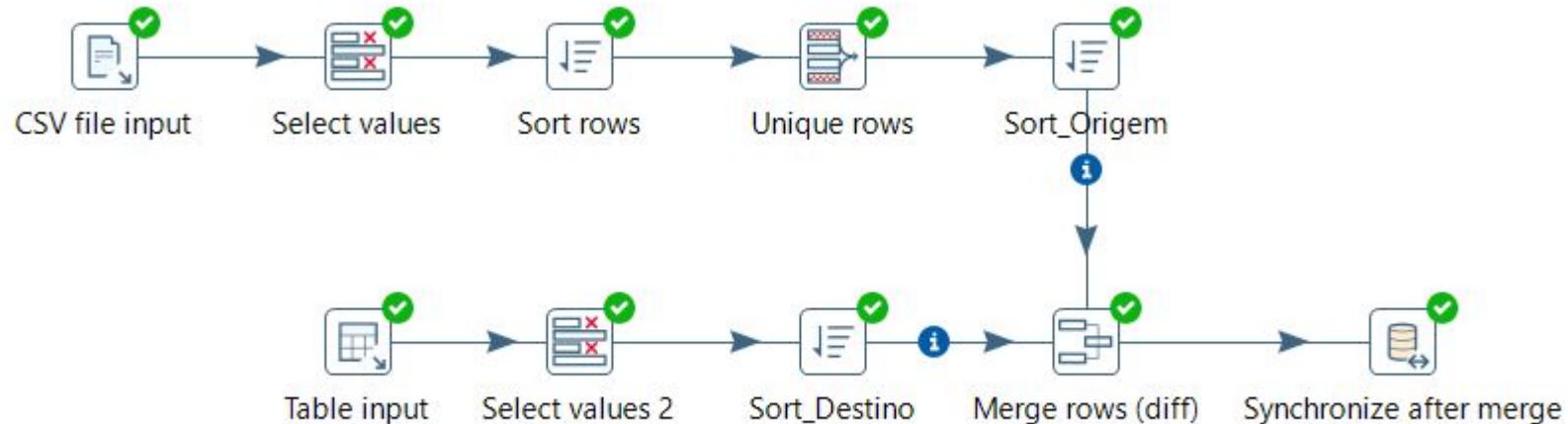
13. Plano de Carga da Dimensão TEMPO



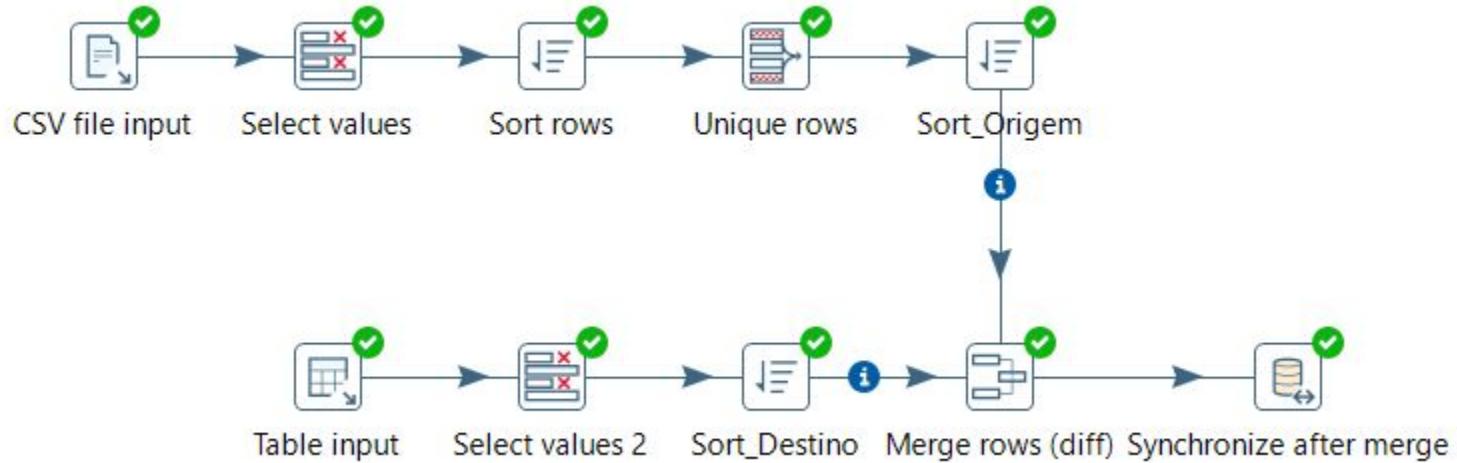
14. Plano de Carga da Dimensão REGIÃO



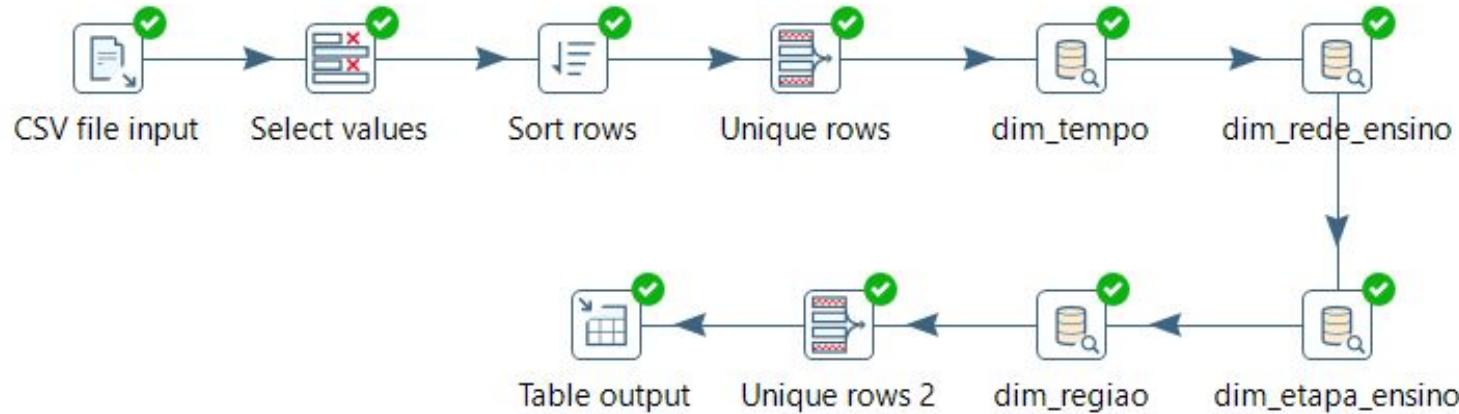
15. Plano de Carga da Dimensão ETAPA ENSINO



16. Plano de Carga da Dimensão REDE ENSINO



17. Plano de Carga da Dimensão ETAPA FATO

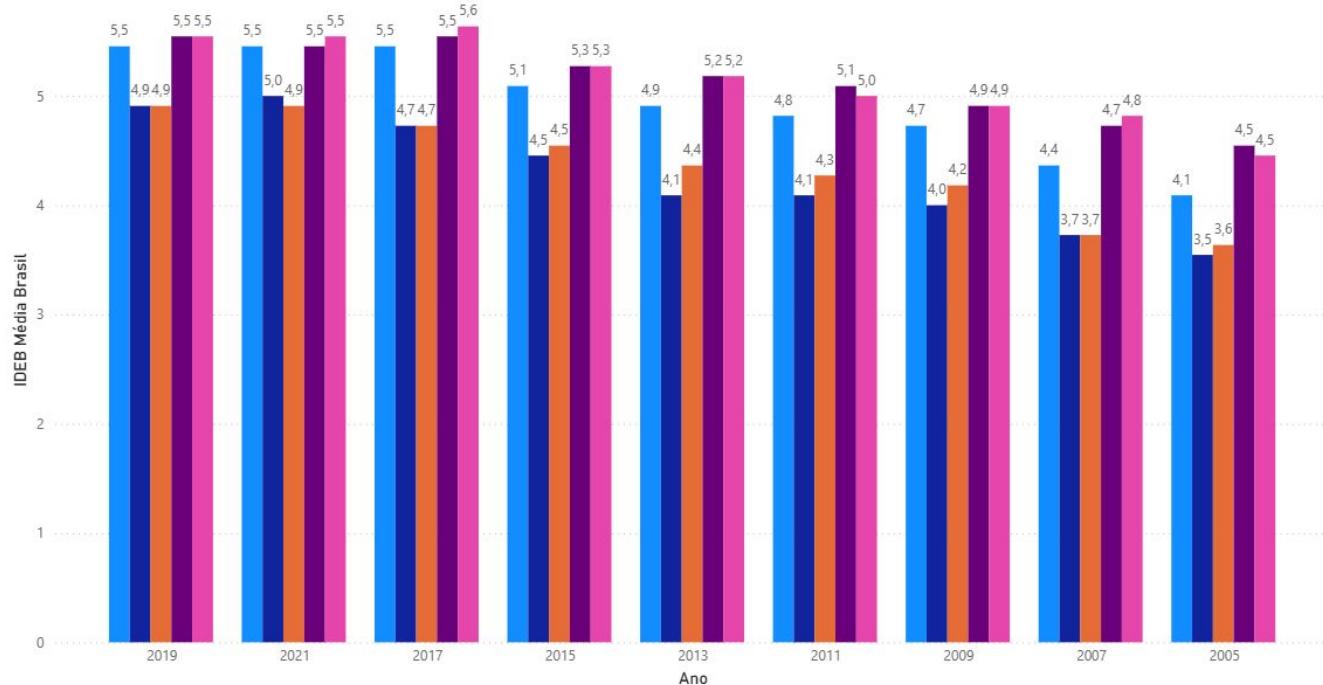


ETAPA 6. APLICAÇÃO OLAP E PAINEL DE BORDO

Consulta OLAP 1

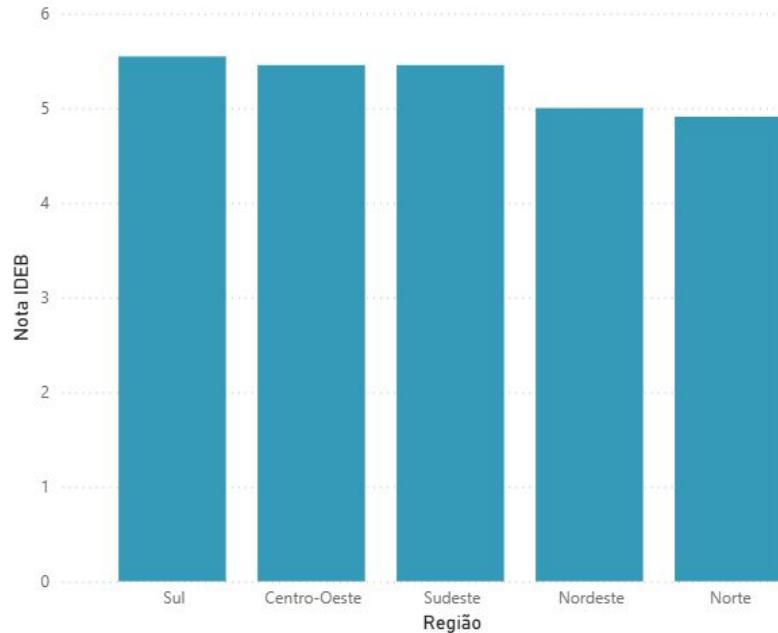
Evolução Histórica do Brasil - (2005-2021)

Região ● Centro-Oeste ● Nordeste ● Norte ● Sudeste ● Sul

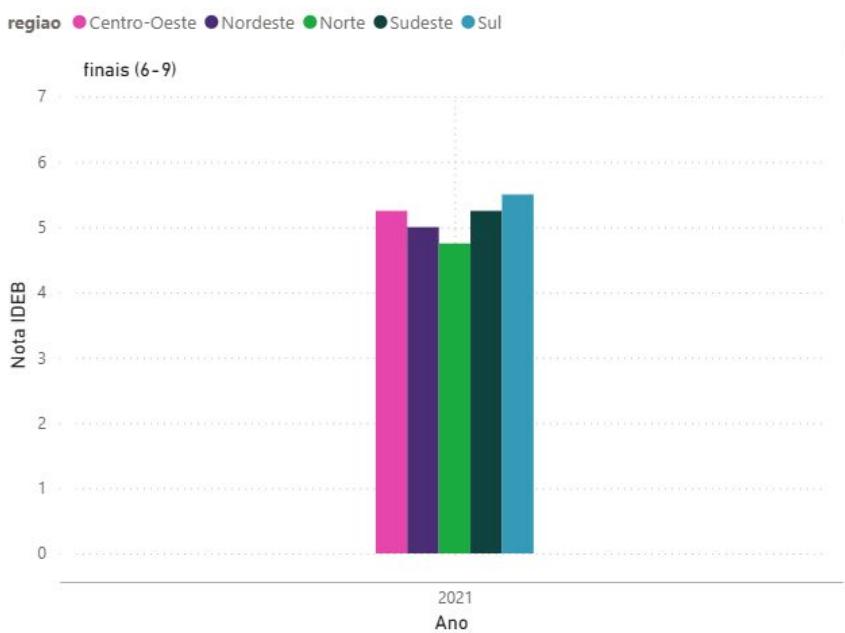


Consulta OLAP 2

Comparativo do IDEB entre as Regiões



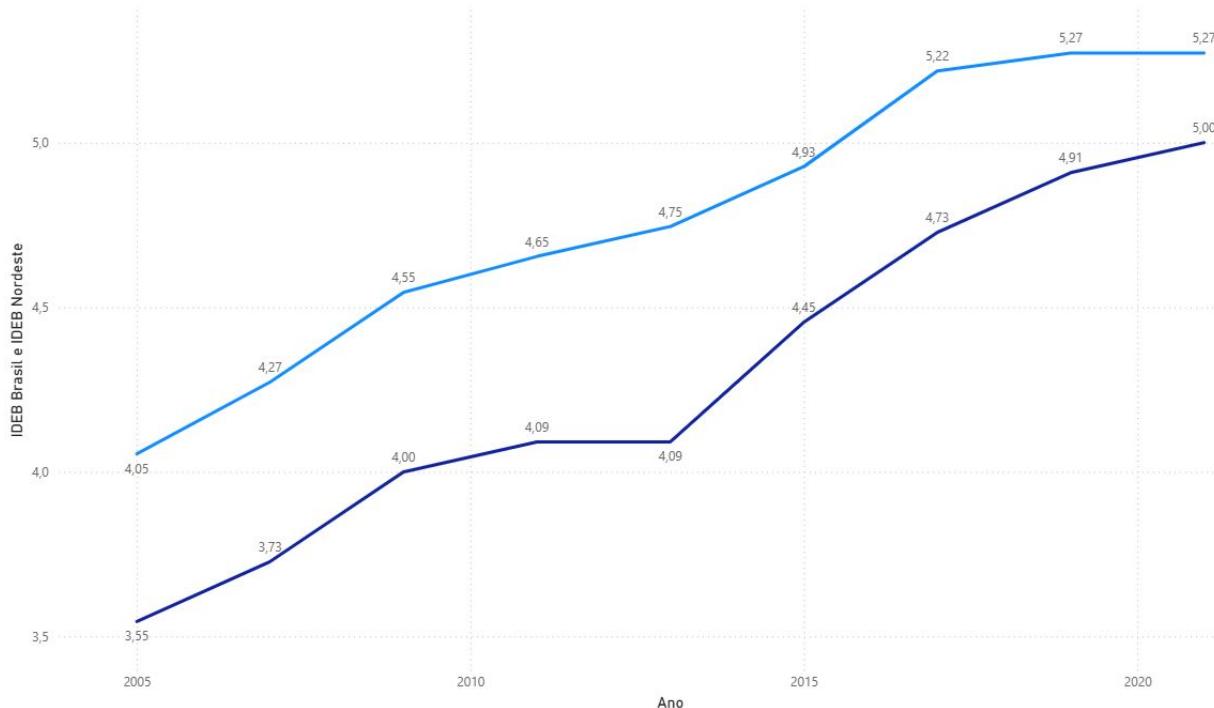
Nota Média do SAEB nas Regiões



Consulta OLAP 3

Comparativo Brasil vs Nordeste

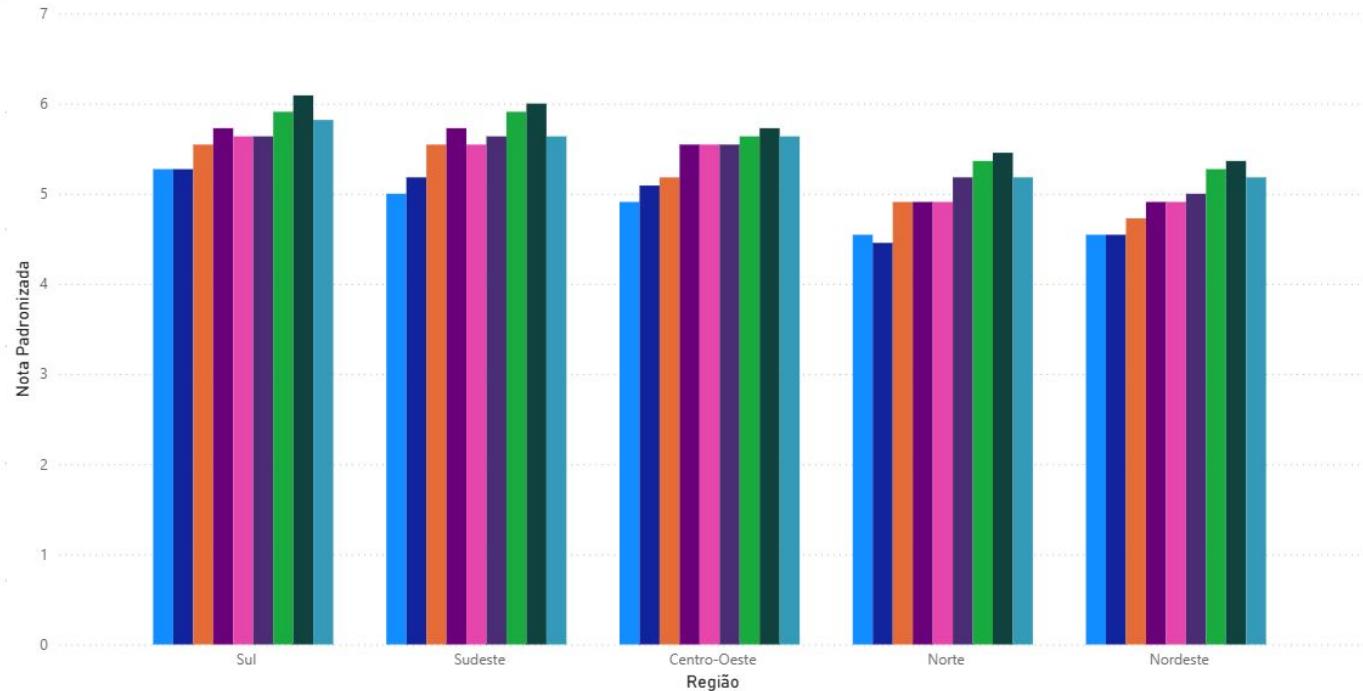
● IDEB Brasil ● IDEB Nordeste



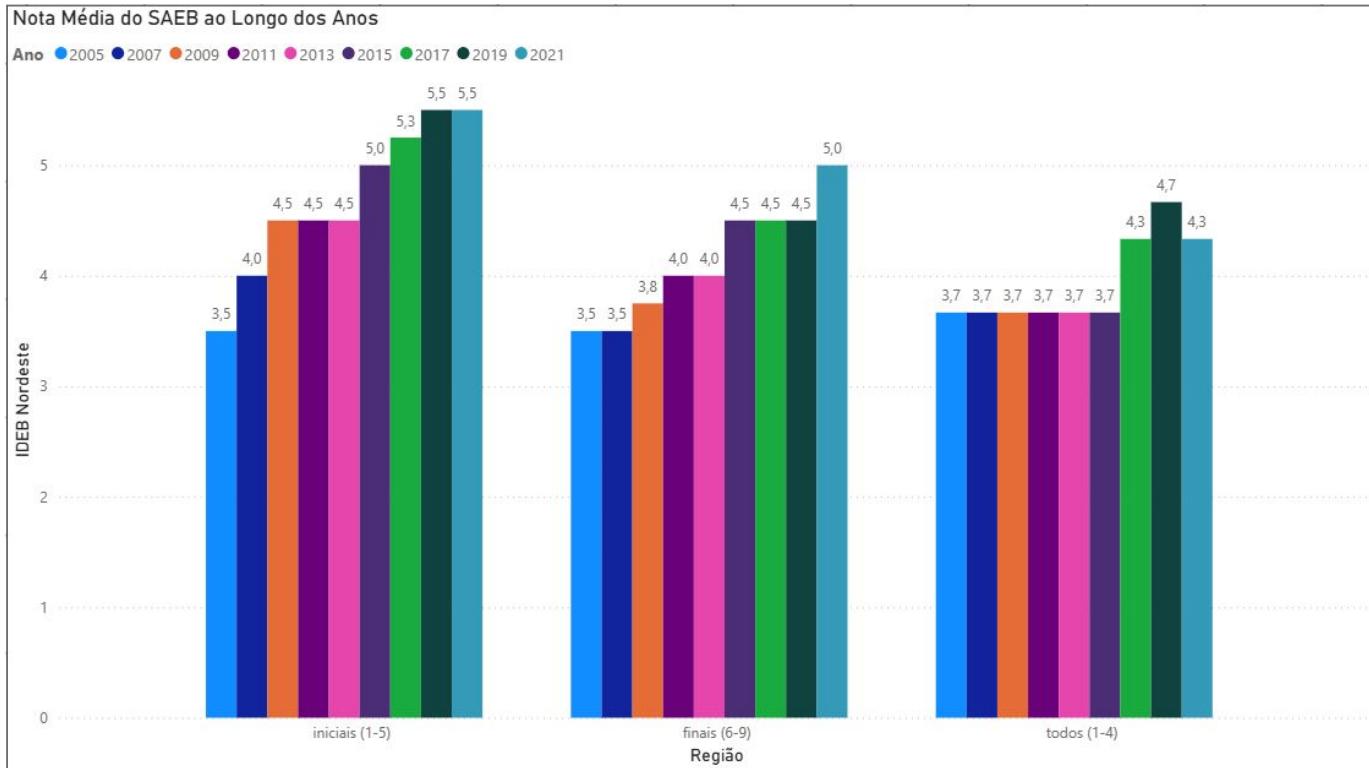
Consulta OLAP 4

Nota Média do SAEB ao Longo dos Anos

Ano ● 2005 ● 2007 ● 2009 ● 2011 ● 2013 ● 2015 ● 2017 ● 2019 ● 2021

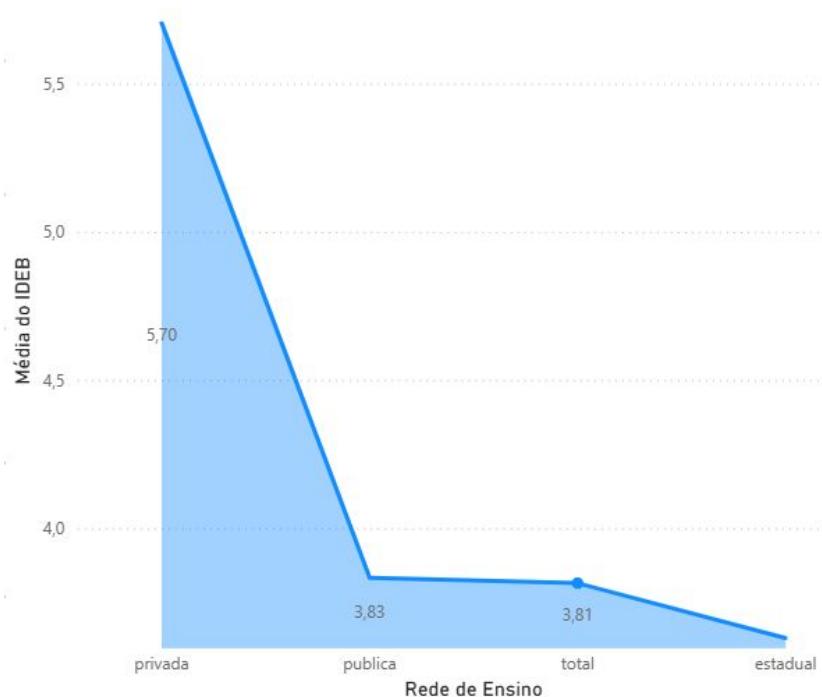


Consulta OLAP 5

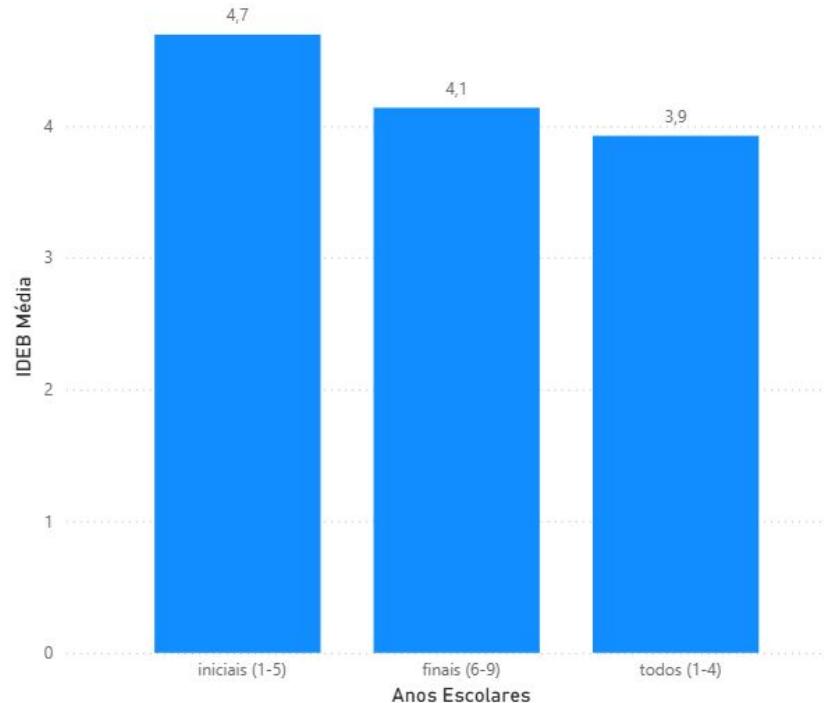


Consulta OLAP 6

Média do IDEB por Rede de Ensino - Nordeste (2005 - 2021)



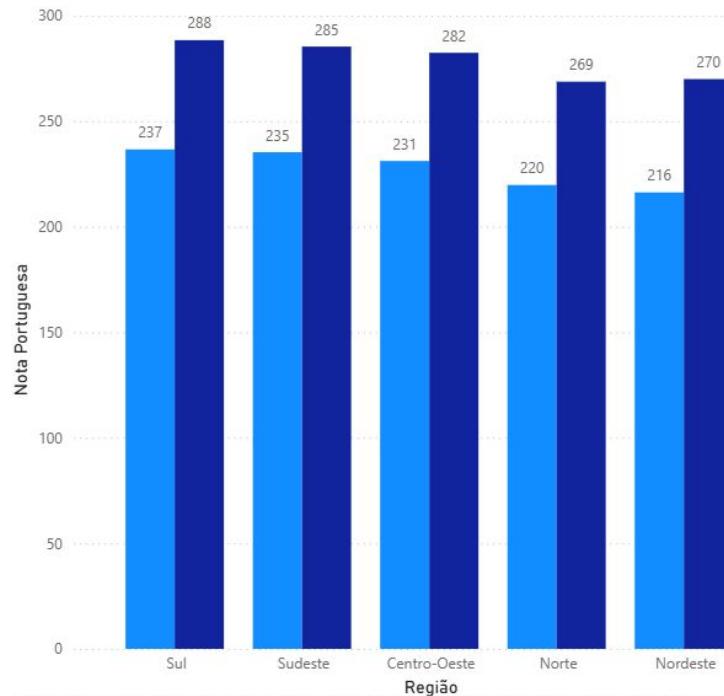
Média do IDEB por Etapa - Nordeste (2005 - 2021)



Consulta OLAP 7

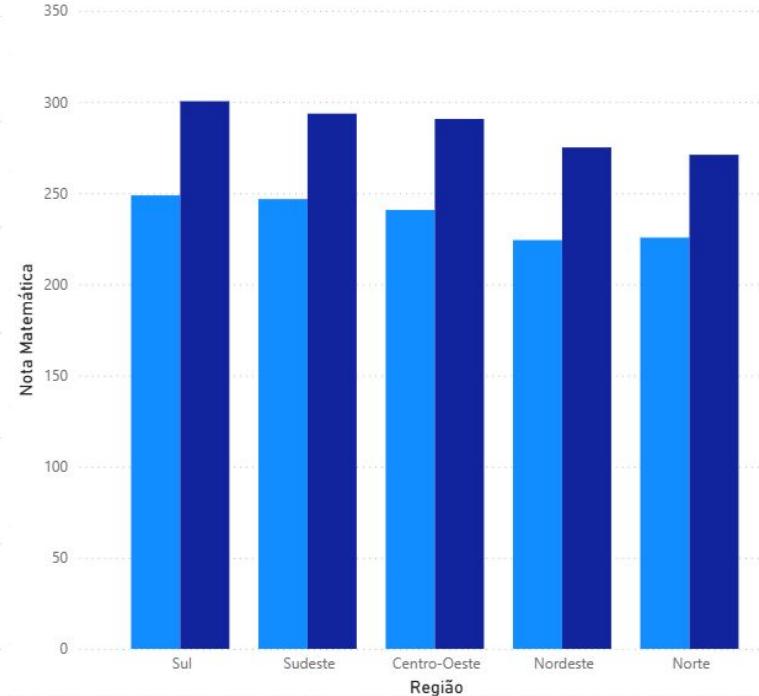
Análise Detalhada da Disciplina de Português por Etapa de Ensino

Etapa ● fundamental ● medio



Análise Detalhada da Disciplina de Matemática por Etapa de Ensino

Etapa ● fundamental ● medio



Referências

1 - Dados Fontes (Abertos).

2- DEMO, P.; SILVA, R. A.; MINAYO, M. C. S. Resultados do IDEB-2019 sugerem avanço no Ensino Médio. Revista Educar Mais, 2021.