

# Estudo de caso **sp**parcerias para vaga de Analista de Dados

Candidato: Rafael Diniz Ramos

## Análise de Desempenho Escolares em Escolas Públicas

Uma Prefeitura tem interesse em avaliar o desempenho escolar de alunos do Ensino Fundamental Regular – Anos Finais, de escolas públicas municipais da capital de São Paulo e de Pernambuco, com foco na comparação entre os anos 2019 e 2023.

Para isso, o analista de dados utilizará o Índice de Desenvolvimento da Educação Básica (IDEB) para análise detalhada. O objetivo é comparar os dados de desempenho entre os municípios, em ambos os períodos, e identificar tendências, melhorias e pontos de atenção.

O banco de dados está disponível no seguinte link: [Resultados — Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep](#)

Primeiramente gostaria de agradecer a oportunidade dada. Vamos lá!

Ao abrir a planilha com os dados e ver que possui um grande cabeçalho com multiníveis, o pedido proposto para análise ter um filtro de município, capital, rede escolar e ano, achei mais prático e rápido excluir linhas e colunas manualmente e restar apenas o que interessa para o caso. Além disso, como faremos ao final gráficos no Power BI, evitaria também problemas ao exportar a planilha.

Abaixo imagens de como ficou nossa base.

In [1]:

```
import pandas as pd
from pandas import DataFrame
```

In [2]:

```
df = pd.read_excel('divulgacao_anos_finais_municipios_2023.xlsx')
```

In [3]:

```
df.head()
```

Out[3]:

	Sigla da UF	Nome do Município	Rede	6º a 9º ano	6º	7º	8º	9º	Indicador de Rendimento (P) - 2019	6º a 9º ano.1	...	9º.1	Indicador de Rendimento (P) - 2023	SAEB - 2019 Matemática	SAEB - 2019 Língua Portuguesa	Nota Média Padronizada (N) - 2019	SAEB - 2023 Matemática	SAEB 2023 Língua Portuguesa
0	RO	Alta Floresta D'Oeste	Estadual	95.2	93.6	95.7	95.7	96.9	0.954601	95.6	...	95.7	0.955774	273.59	268.37	5.699333	258.13	2
1	RO	Alta Floresta D'Oeste	Municipal	88.9	91.1	86.2	86.6	91.6	0.888055	87.6	...	96.3	0.875571	255.87	252.15	5.133667	238.44	2
2	RO	Alta Floresta D'Oeste	Pública	93.2	92.8	92.9	92.6	95.3	0.933872	93.2	...	95.9	0.931733	270.02	265.1	5.585333	254.11	2
3	RO	Ariquemes	Estadual	92	90.6	91.5	92.9	93.5	0.921108	94.4	...	91.4	0.945972	267.39	266.19	5.559667	252.55	2
4	RO	Ariquemes	Municipal	89.2	90	84.4	86.9	94.3	0.887481	91.9	...	95.1	0.921178	247.56	251.95	4.991833	237.33	2

5 rows × 23 columns

In [13]:

```
df.columns
```

Out[13]:

```
Index(['Sigla da UF', 'Nome do Município', 'Rede', '6º a 9º ano', '6º', '7º', '8º', '9º', 'Indicador de Rendimento (P) - 2019', '6º a 9º ano.1', '6º.1', '7º.1', '8º.1', '9º.1', 'Indicador de Rendimento (P) - 2023', 'SAEB - 2019 Matemática', 'SAEB - 2019 Língua Portuguesa', 'Nota Média Padronizada (N) - 2019', 'SAEB - 2023 Matemática', 'SAEB - 2023 Língua Portuguesa', 'Nota Média Padronizada (N) - 2023', 'IDEB\2019\n(N x P)', 'IDEB\2023\n(N x P)'], dtype='object')
```

Abaixo eu reordeno os índices das colunas, pois quero por as mesmas na ordem cronológica (2019 e 2023), para depois cortar e fazer duas tabelas dividas pelo período.

In [18]:

```
df_reor = df.iloc[:, [0, 1, 2, 3, 4, 5, 6, 7, 8, 15, 16, 17, 21, 9, 10, 11, 12, 13, 14, 18, 19, 20, 22]]
df_reor
```

Out[18]:

	Sigla da UF	Nome do Município	Rede	6 a 9 ano	6º ano	7º ano	8º ano	9º ano	Indicador de Rendimento (P) - 2019	SAEB - 2019 Matemática	...	6 a 9 ano.1	6º ano.1	7º ano.1	8º ano.1	9º ano.1	Indicador de Rendimento (P) - 2023	SAEB - 2023 Matemática	S L
0	RO	Alta Floresta D'Oeste	Estadual	95.2	93.6	95.7	95.7	96.9	0.954601	273.59	...	95.6	97	93.2	96.5	95.7	0.955774	258.13	2
1	RO	Alta Floresta D'Oeste	Municipal	88.9	91.1	86.2	86.6	91.6	0.888055	255.87	...	87.6	83.6	88.8	82.8	96.3	0.875571	238.44	2
2	RO	Alta Floresta D'Oeste	Pública	93.2	92.8	92.9	92.6	95.3	0.933872	270.02	...	93.2	92.8	91.9	92.2	95.9	0.931733	254.11	2
3	RO	Ariquemes	Estadual	92	90.6	91.5	92.9	93.5	0.921108	267.39	...	94.4	97.4	95.8	94	91.4	0.945972	252.55	2
4	RO	Ariquemes	Municipal	89.2	90	84.4	86.9	94.3	0.887481	247.56	...	91.9	91.5	91.4	90.6	95.1	0.921178	237.33	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14396	GO	Vila Rica	Estadual	98.2	96.9	96.8	100	98.8	0.981067	268.96	...	99.4	98.9	99	100	100	0.994722	268.26	2

Definindo, dividindo e salvando as tabelas por ano.

```
In [19]: coluna_corte = df_reor.columns.get_loc('IDEB\n2019\n(N x P)') + 1

In [20]: df_2019 = df_reor.iloc[:, :coluna_corte]
df_2023 = df_reor.iloc[:, coluna_corte:]

In [21]: df_2019.to_excel('df_2019.xlsx', index=False)
df_2023.to_excel('df_2023.xlsx', index=False)
```

Renomeando colunas

```
In [5]: df_2019 = df_2019.rename(columns={'IDEB\n2019\n(N x P)': 'IDEB 2019 (N x P)'})

In [7]: df_2023 = df_2023.rename(columns={
    '6 a 9 ano.1': '6 a 9 ano',
    '6 ano.1': '6 ano',
    '7 ano.1': '7 ano',
    '8 ano.1': '8 ano',
    '9 ano.1': '9 ano',
    'IDEB\n2023\n(N x P)': 'IDEB 2023 (N x P)'
})
```

Mostrando as tabelas divididas por ano

2019

```
In [13]: df_2019 = pd.read_excel('df_2019.xlsx')
df_2023 = pd.read_excel('df_2023.xlsx')

In [22]: df_2019

Out[22]:
```

	Sigla da UF	Nome do Município	Rede	6 a 9 ano	6 ano	7 ano	8 ano	9 ano	Indicador de Rendimento (Pi) - 2019	SAEB - 2019 Matemática	SAEB - 2019 Língua Portuguesa	Nota Média Padronizada (N) - 2019	IDEB 2019 (N x P)
0	RO	Alta Floresta D'Oeste	Estadual	95.2	93.6	95.7	95.7	96.9	0.954601	273.59	268.37	5.699333	5.4
1	RO	Alta Floresta D'Oeste	Municipal	88.9	91.1	86.2	86.6	91.6	0.888055	255.87	252.15	5.133667	4.6
2	RO	Alta Floresta D'Oeste	Pública	93.2	92.8	92.9	92.6	95.3	0.933872	270.02	265.1	5.585333	5.2
3	RO	Ariquemes	Estadual	92	90.6	91.5	92.9	93.5	0.921108	267.39	266.19	5.559667	5.1
4	RO	Ariquemes	Municipal	89.2	90	84.4	86.9	94.3	0.887481	247.56	251.95	4.991833	4.4
...	...	...	...	...	...	...	...	...	...	...	...	...	...
14396	GO	Vila Propício	Estadual	98.2	96.9	96.8	100	98.8	0.981067	268.96	269.46	5.640333	5.5
14397	GO	Vila Propício	Pública	98.2	96.9	96.8	100	98.8	0.981067	268.96	269.46	5.640333	5.5
14398	DF	Brasília	Federal	96.3	94.5	98.2	96.9	95.3	0.962038	ND	ND	-	-
14399	DF	Brasília	Estadual	89.1	96.2	81.5	96.2	84	0.889586	257.14	254.28	5.190333	4.6
14400	DF	Brasília	Pública	89.2	96.2	81.6	96.2	84.1	0.890164	258.05	254.85	5.215	4.6

14401 rows x 13 columns

2023

In [33]:

df\_2023

Out[33]:

	Sigla da UF	Nome do Município	Rede	6 a 9 ano	6 ano	7 ano	8 ano	9 ano	Indicador de Rendimento (P) - 2023	SAEB - 2023 Matemática	SAEB - 2023 Língua Portuguesa	Nota Média Padronizada (N) - 2023	IDEB 2023 (N x P)
0	RO	Alta Floresta D'Oeste	Estadual	95.6	97	93.2	96.5	95.7	0.955774	258.13	249.46	5.1265	4.9
1	RO	Alta Floresta D'Oeste	Municipal	87.6	83.6	88.8	82.8	96.3	0.875571	238.44	233.56	4.533333	4
2	RO	Alta Floresta D'Oeste	Pública	93.2	92.8	91.9	92.2	95.9	0.931733	254.11	246.21	5.005333	4.7
3	RO	Ariquemes	Estadual	94.4	97.4	95.8	94	91.4	0.945972	252.55	252.21	5.079333	4.8
4	RO	Ariquemes	Municipal	91.9	91.5	91.4	90.6	95.1	0.921178	237.33	246.46	4.729833	4.4
...	...	...	...	...	...	...	...	...	...	...	...	...	...
14396	GO	Vila Propício	Estadual	99.4	98.9	99	100	100	0.994722	268.26	276.37	5.743833	5.7
14397	GO	Vila Propício	Pública	99.4	98.9	99	100	100	0.994722	268.26	276.37	5.743833	5.7
14398	DF	Brasília	Federal	98.3	98.1	97.3	99.2	98.4	0.982453	328.46	302.97	7.1905	7.1
14399	DF	Brasília	Estadual	92.4	97.1	87.9	96.1	88.8	0.922886	245.87	250.11	4.933	4.6
14400	DF	Brasília	Pública	92.5	97.1	88	96.1	88.9	0.923431	247.08	250.89	4.966167	4.6

14401 rows × 13 columns

Filtrando por UF, Município e Rede

2019

In [29]:

df\_2019\_filtrado = df\_2019[(df['Sigla da UF'].isin(['SP', 'PE'])) & (df\_2019['Nome do Município'].isin(['São Paulo', 'Recife'])) & (df\_2019['Rede'] == 'Municipal')]

In [30]:

df\_2019\_filtrado

Out[30]:

	Sigla da UF	Nome do Município	Rede	6 a 9 ano	6 ano	7 ano	8 ano	9 ano	Indicador de Rendimento (P) - 2019	SAEB - 2019 Matemática	SAEB - 2019 Língua Portuguesa	Nota Média Padronizada (N) - 2019	IDEB 2019 (N x P)
4357	PE	Recife	Municipal	94.3	92.2	92.9	96	97.6	0.946235	248.83	246.59	4.923667	4.7
10054	SP	São Paulo	Municipal	94.6	94.2	94.2	93.9	96.4	0.946645	251.49	252.97	5.074333	4.8

2023

In [31]:

df\_2023\_filtrado = df\_2023[(df['Sigla da UF'].isin(['SP', 'PE'])) & (df\_2023['Nome do Município'].isin(['São Paulo', 'Recife'])) & (df\_2023['Rede'] == 'Municipal')]

In [32]:

df\_2023\_filtrado

Out[32]:

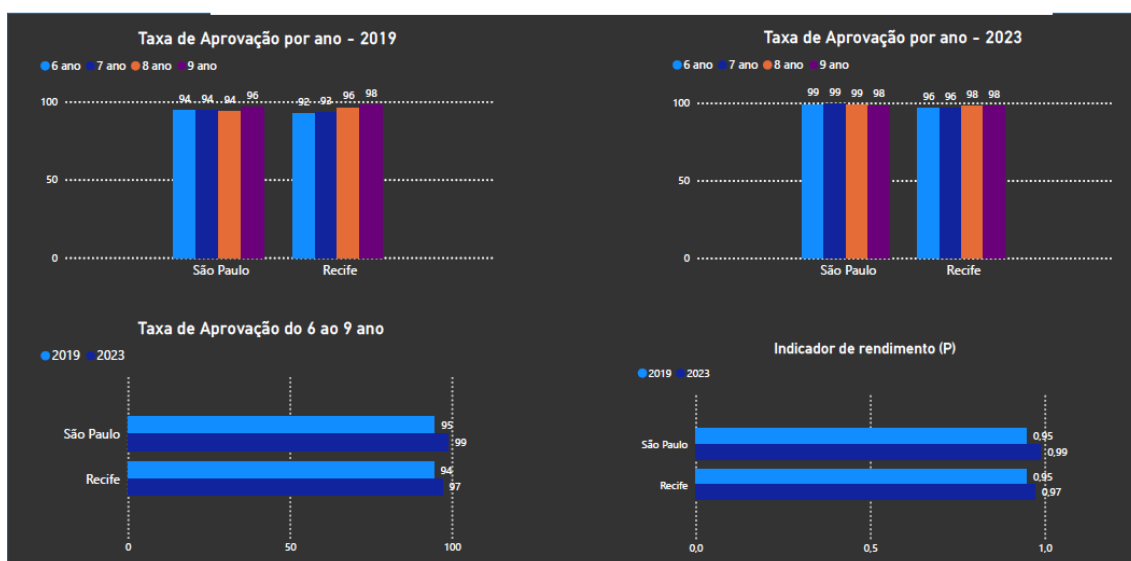
	Sigla da UF	Nome do Município	Rede	6 a 9 ano	6 ano	7 ano	8 ano	9 ano	Indicador de Rendimento (P) - 2023	SAEB - 2023 Matemática	SAEB - 2023 Língua Portuguesa	Nota Média Padronizada (N) - 2023	IDEB 2023 (N x P)
4357	PE	Recife	Municipal	97	96.4	96.3	97.7	98.1	0.971186	244.73	253.89	4.977	4.8
10054	SP	São Paulo	Municipal	98.8	99	99.2	98.7	98.3	0.987988	244.77	247.08	4.864167	4.8

# Construção dos gráficos no Power BI



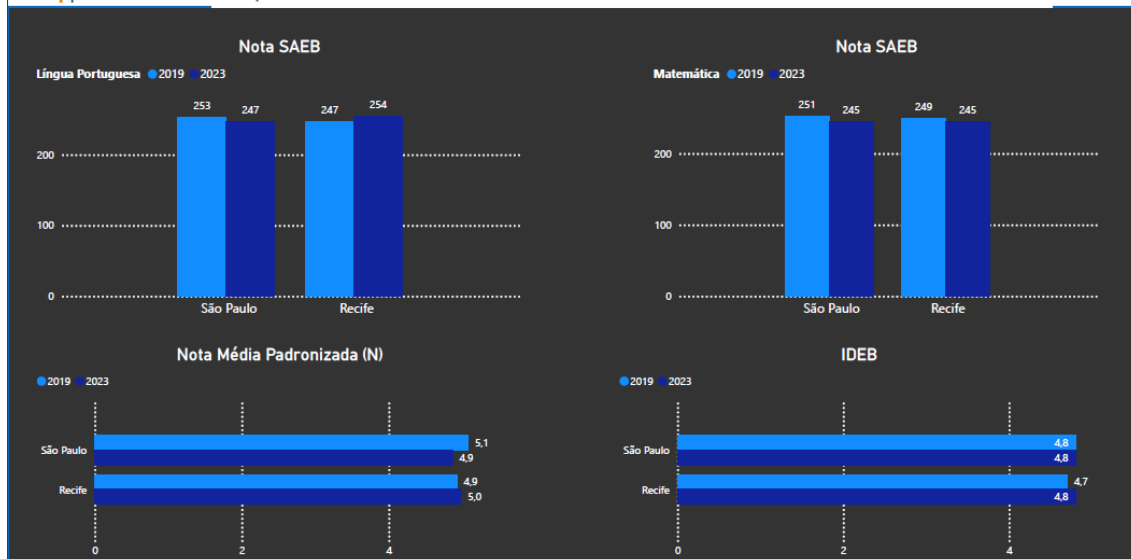
## Estudo de caso: Análise de Desempenho Escolares em Escolas Públicas

Comparativos de índices escolares da cidade de São Paulo e Recife entre 2019 e 2023



## Estudo de caso: Análise de Desempenho Escolares em Escolas Públicas

Comparativos de índices escolares da cidade de São Paulo e Recife entre 2019 e 2023



Pontos a destacar sobre a análise:

1. A taxa de aprovação do 6º ao 9º ano cresceu em ambas as cidades, com São Paulo tendo um crescimento de um ponto a mais do que Recife, no período de 2019 a 2023.
2. O indicador de rendimento também evoluiu em ambas as cidades, novamente com São Paulo ter ficado a frente por um ponto apenas, no período de 2019 a 2023.
3. A nota do SAEB em Língua Portuguesa, diminuiu 6 pontos em São Paulo e aumentou 7 pontos em Recife.
4. A nota do SAEB em Matemática, caiu em as ambas as cidades, sendo 6 pontos em São Paulo e 4 pontos em Recife. Aqui vale uma reflexão sobre o que pode ser feito para que se recupere o desempenho na disciplina.
5. A nota média padronizada (N) caiu um pouco nas duas cidades analisadas.
6. O IDEB se manteve no mesmo índice em São Paulo e em Recife subiu 0,1 ponto.

Este documento e o arquivo do Power BI estão disponíveis em meu Github

[https://github.com/devrafael26/case\\_spparcerias](https://github.com/devrafael26/case_spparcerias).