

Projeto 3: Análise de Desigualdade Educacional no Brasil

Esse notebook servirá para anotar as ideias, as análises aleatorias feitas e demais coisas.

Em outro notebook terá a organização bem feita das análises

Tema:

Vamos usar dados do INEP (Ideb) ou Censo Escolar para analisar desigualdades educacionais entre diferentes regiões do Brasil. O foco será comparar indicadores como taxa de aprovação, evasão, número de alunos por turma e IDEB em diferentes estados, anos e redes de ensino (pública x privada).

Fonte de Dados:

Para os dados, utilizei os microdados dos dados abertos do Censo Escolar 2024, que podem ser conferidas logo abaixo:

- [Microdados do Censo Escolar](#)

Uma mudança

A base de dados do Censo Escolar é muito extensa para eu trabalhar como um iniciante. Eu decidi trabalhar apenas com um estado inicialmente, e posteriormente ir trabalhando com os outros estados, captando pequenas mudanças se necessário.

Ao observar os dados, e também o dicionário de dados, foi perceptível que existe muita informação ao qual da para ser trabalhada com apenas um estado, e assim conseguirei atingir meus objetivos iniciais. Outro motivo também foi porque assim também terei uma facilidade no entendimento do contexto local dos estados.

Irei então trabalhar com o estado de **Pernambuco**

```
import pandas as pd
```

```
# O arquivo que antes estava nesse local foi apagado, pois pesava muito
# df =
pd.read_csv("microdados_censo_escolar_2024/dados/microdados_ed_basica_2024.csv",
sep=';', encoding='latin1')
# df_estado = df[df['SG_UF'] == 'PE']
```

```
# Salvando o DataFrame só com as informações do estado de 'Pernambuco'
df_estado.to_csv('microdados_censo_PE_2024.csv', index=False)
```

```
# Abaixo está o arquivo csv salvo após filtrar apenas para um único estado
df = pd.read_csv("microdados_censo_escolar_2024/dados/microdados_censo_PE_2024.csv")
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9936 entries, 0 to 9935
Columns: 426 entries, NU_ANO_CENSO to QT_TUR_MED_INT
dtypes: float64(396), int64(14), object(16)
memory usage: 32.3+ MB
```

```
df.columns
```

```
Index(['NU_ANO_CENSO', 'NO_REGIAO', 'CO_REGIAO', 'NO_UF', 'SG_UF', 'CO_UF',
       'NO_MUNICIPIO', 'CO_MUNICIPIO', 'NO_REGIAO_GEOG_INTERM',
       'CO_REGIAO_GEOG_INTERM',
       ...,
       'QT_TUR_BAS_D', 'QT_TUR_BAS_N', 'QT_TUR_BAS_EAD', 'QT_TUR_INF_INT',
       'QT_TUR_INF_CRE_INT', 'QT_TUR_INF_PRE_INT', 'QT_TUR_FUND_INT',
       'QT_TUR_FUND_AI_INT', 'QT_TUR_FUND_AF_INT', 'QT_TUR_MED_INT'],
      dtype='object', length=426)
```

```
df.isnull().sum()
```

```
NU_ANO_CENSO      0
NO_REGIAO         0
CO_REGIAO         0
NO_UF             0
SG_UF             0
...
QT_TUR_INF_PRE_INT 1943
QT_TUR_FUND_INT    1943
QT_TUR_FUND_AI_INT 1943
QT_TUR_FUND_AF_INT 1943
QT_TUR_MED_INT     1943
Length: 426, dtype: int64
```

```
df.head()
```

border="1" class="dataframe">

	N U A N O C E N S O	N O R E G I A O	C O R E G I A O	N O U F	S G U F	C O U F	N O M U N I C I P I O	C O M U N I C I P I O	N O R E G I A O G E O G I N T E R M	C O R E G I A O G E O G I N T E R M	...	Q T U R B A S D	Q T U R B A S N	Q T U R B A S E A D	Q T U R I N F I N T	Q T U R I N F C R E I N T	Q T U R I N F P R E I N T	Q T U R F U N D I N T	Q T U R F U N D A I N T	Q T U R F U N D A F I N T	Q T U R M E D I N T
0	20 24	No rd este	2	Pe rn a m bu co	PE	26	Ab re u e Li m a	26 00 05 4	Re cif e	26 01	...	Na N	Na N	Na N	Na N	Na N	Na N	Na N	Na N	Na N	Na N
1	20 24	No rd este	2	Pe rn a m bu co	PE	26	Ab re u e Li m a	26 00 05 4	Re cif e	26 01	...	18 .0	1. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0
2	20 24	No rd este	2	Pe rn a m bu co	PE	26	Ab re u e Li m a	26 00 05 4	Re cif e	26 01	...	12 .0	2. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0
3	20 24	No rd este	2	Pe rn a m bu co	PE	26	Ab re u e Li m a	26 00 05 4	Re cif e	26 01	...	15 .0	9. 0	0. 0	0. 0	0. 0	0. 0	3. 0	0. 0	3. 0	0. 0
4	20 24	No rd este	2	Pe rn a m bu co	PE	26	Ab re u e Li m a	26 00 05 4	Re cif e	26 01	...	Na N	Na N	Na N	Na N	Na N	Na N	Na N	Na N	Na N	Na N

5 rows × 426 columns

```
df.tail()
```

border="1" class="dataframe">

N_U_A_N_O_C_E_N_S_O		N_O_R_E_G_I_A_O		C_O_R_E_G_I_A_O		N_O_U_F		S_G_U_F		C_O_U_F		N_O_M_U_N_I_C_I_P_I_O		C_O_M_U_N_I_C_I_P_I_O		N_O_R_E_G_I_A_O_G_I_N_T_E_R_M		C_O_R_E_G_I_A_O_G_I_N_T_E_R_M		...		Q_T_T_U_R_B_A_S_D		Q_T_T_U_R_B_A_S_N		Q_T_T_U_R_B_A_S_E_A_D		Q_T_T_U_R_I_N_F_I_N_T		Q_T_T_U_R_I_N_F_C_R_E_I_N_T		Q_T_T_U_R_I_N_F_P_R_E_I_N_T		Q_T_T_U_R_I_N_F_U_N_D_I_N_T		Q_T_T_U_R_I_N_F_U_N_D_A_I_N_T		Q_T_T_U_R_I_N_F_U_N_D_A_F_I_N_T		Q_T_T_U_R_I_N_F_U_N_D_I_N_T	
9931	2024	Nordeste	2	Pernambuco	PE	26	Xexéu	2616506	Recife	2601	...	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9932	2024	Nordeste	2	Pernambuco	PE	26	Xexéu	2616506	Recife	2601	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9933	2024	Nordeste	2	Pernambuco	PE	26	Xexéu	2616506	Recife	2601	...	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
9934	2024	Nordeste	2	Pernambuco	PE	26	Xexéu	2616506	Recife	2601	...	11.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
9935	2024	Nordeste	2	Pernambuco	PE	26	Xexéu	2616506	Recife	2601	...	12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

5 rows × 426 columns

Importante

Ao observar o dicionário de dados (está em Anexo I), foi visto que não tem uma maneira de reduzir a quantidade de colunas do dataframe, que se encontram em 426.

Com isso em mente, irei tentar trabalhar da melhor maneira possível para obter uma análise de boa qualidade para esse meu pequeno projeto. Irei sempre estar utilizando o dicionário de dados para conseguir identificar o que determinada coluna me informa, ou até mesmo para encontrar mais facilmente alguma informação de coluna para utilizar.

```
# Todas as colunas do dataframe
#df.columns.tolist()
```

Algumas Ideias...



Algumas perguntas que estão surgindo para a análise:

- Qual a taxa de aprovação média por município?
- A infraestrutura das escolas (biblioteca, laboratório, internet) influencia nos resultados?
- Quantas escolas por tipo de rede (pública, privada)?
- Qual o perfil dos docentes?

Vou separar por categorias, acredito que fica melhor.



Perfil das Escolas

- Qual a quantidade total de escolas por rede (pública x privada)?
- Qual o percentual de escolas que possuem biblioteca, laboratório de informática, quadra esportiva ou acesso à internet?
- Quais são os municípios com maior número de escolas?
- Qual o percentual de escolas urbanas x rurais?
- Qual a média de alunos por escola?
- Qual a distribuição das escolas por etapas de ensino (fundamental, médio, EJA, etc.)?



Perfil dos Alunos

- Quantos alunos estão matriculados no estado por rede de ensino?
- Qual a média de alunos por turma (ou sala)?
- Qual a proporção de alunos na zona urbana x zona rural?
- Qual o percentual de alunos com necessidades especiais atendidos nas escolas?
- Quantos alunos por município? Existe concentração em alguns polos?
- Existe diferença de matrícula por sexo?



Perfil dos Docentes

- Qual a quantidade total de docentes no estado?
- Qual o percentual de docentes com ensino superior completo?

- Qual a média de alunos por docente?
- Quais são os municípios com mais professores?
- Existem diferenças na formação dos docentes entre escolas públicas e privadas?

Infraestrutura Escolar x Indicadores

b- Existe relação entre a presença de laboratório de informática e o número médio de alunos?

- Qual a infraestrutura mais presente nas escolas públicas? E nas privadas?
- Infraestrutura varia muito entre zonas urbana e rural?

Comparativos

- Qual a diferença de estrutura e número de alunos entre escolas urbanas e rurais?
- Qual a diferença entre as redes pública e privada?
- Como está a distribuição das escolas por município em relação à população (ex: escolas per capita)?

Possíveis Cruzes de Dados (Para aprofundar)

- Municípios com mais escolas públicas possuem também maior número de docentes qualificados?
- Escolas públicas em áreas urbanas têm infraestrutura melhor que em áreas rurais?
- Escolas que oferecem mais etapas de ensino concentram mais alunos?

Análises Temporais (Se tiver mais anos)

- O número de matrículas cresceu ou diminuiu ao longo dos anos?
- Como a pandemia afetou as matrículas ou número de docentes?
- A infraestrutura das escolas melhorou nos últimos anos?



Sugestões de Análises para o Projeto

1 Panorama das Escolas no Estado

- Quantas escolas existem no estado por rede de ensino (pública/privada)?
- Quantas escolas têm biblioteca, laboratório, internet?
- Distribuição do número de alunos por escola.

Insight esperado: “No estado de SP, apenas 40% das escolas públicas possuem laboratório de informática.”

2 Infraestrutura x Rede de Ensino

- Qual a diferença de infraestrutura entre escolas públicas e privadas?
- Existe diferença significativa no acesso à internet, quadra, biblioteca?
- Gráficos sugeridos:
 - Barras comparativas
 - Heatmap de correlação por rede

3 Número de Alunos por Município

- Quais são os 10 municípios com maior número de alunos matriculados?
- Existe relação entre o tamanho do município (número de escolas) e o número médio de alunos por escola?

Possível insight: “Municípios menores têm turmas mais cheias em escolas públicas.”

4 Docentes e Qualificação

- Qual a proporção de docentes com formação superior por rede?
- Há diferença na formação de professores da rede pública e privada?
- Gráfico sugerido: Barras empilhadas ou porcentagem

5 Análise Temporal (se tiver mais de um ano)

- Como evoluiu o número de matrículas nos últimos anos?
- Houve queda de matrícula na pandemia?

6 Geolocalização (Extra)

- Se quiser ousar, pode mapear as escolas no estado usando as coordenadas (se vierem na base).
- Ou plotar por município.

```
# Analise aleatória
colunas_desejadas = ['NO_MUNICIPIO', 'NO_ENTIDADE', 'IN_AGUA_POTAVEL',
                    'IN_AGUA_REDE_PUBLICA', 'IN_AGUA_POCO_ARTESIANO', 'IN_AGUA_CACIMBA',
                    'IN_AGUA_FONTE_RIO', 'IN_AGUA_INEXISTENTE', 'IN_AGUA_CARRO_PIPA']
```

```
# Visualizar as colunas sobre a agua das escolas
df_copia1 = df[colunas_desejadas]
```

```
df_copia1[df_copia1['NO_MUNICIPIO'] == 'Itambé'].head()
```

border="1" class="dataframe">

	NO_MUNICIPIO	NO_ENTIDADE	IN_AGUA_PO_TAVEL	IN_AGUA_REDE_PUBLICA	IN_AGUA_PO_CO_AR_TESIANO	IN_AGUA_CA_CIMBA	IN_AGUA_FONTE_RIO	IN_AGUA_INEXISTENTE	IN_AGUA_CARRRO_PI_PA
3819	Itambé	ESCOLA DE REFERENCIA EM ENSINO MEDIO FREI ORLANDO	1.0	1.0	0.0	0.0	0.0	0.0	0.0
3820	Itambé	COLEGIO MUNICIPAL PROFESSOR NIVALDO XAVIER DE ...	1.0	1.0	0.0	1.0	0.0	0.0	0.0
3821	Itambé	ESCOLA ARRUDA CAMARUA	1.0	1.0	0.0	1.0	0.0	0.0	0.0
3822	Itambé	ESCOLA MUL ALVARO VELOSO BORBA	1.0	0.0	0.0	1.0	0.0	0.0	0.0
3823	Itambé	ESC MUL ANTONIO GUEDES CORREIA GONDIM	1.0	0.0	0.0	1.0	0.0	0.0	0.0

Perfil das Escolas

- Qual a quantidade total de escolas por rede (pública x privada)?
- Qual o percentual de escolas que possuem biblioteca, laboratório de informática, quadra esportiva ou acesso à internet?

- Quais são os municípios com maior número de escolas?
- Qual o percentual de escolas urbanas x rurais?
- Qual a média de alunos por escola?
- Qual a distribuição das escolas por etapas de ensino (fundamental, médio, EJA, etc.)?

```
# Quantas escolas estão em situação de não funcionamento?
df['TP_SITUACAO_FUNCIONAMENTO'].value_counts()
```

```
TP_SITUACAO_FUNCIONAMENTO
1      8013
2     1809
3      114
Name: count, dtype: int64
```

- Existem 8013 escolas **Em atividade** no estado de Pernambuco;
- Ao momento são 1809 escolas **paralisadas**; e
- 114 escolas foram **extintas** desde o último censo.

```
colunas_desejadas_vinculo = ['NO_MUNICIPIO', 'NO_ENTIDADE',
                             'IN_VINCULO_SECRETARIA_EDUCACAO',
                             'IN_VINCULO_SEGURANCA_PUBLICA', 'IN_VINCULO_SECRETARIA_SAUDE',
                             'IN_VINCULO_OUTRO_ORGAO', 'IN_PODER_PUBLICO_PARCERIA',
                             'TP_PODER_PUBLICO_PARCERIA']
```

```
df_vinculo = df[colunas_desejadas_vinculo]
```

```
df_vinculo.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9936 entries, 0 to 9935
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   NO_MUNICIPIO                          9936 non-null   object
1   NO_ENTIDADE                           9936 non-null   object
2   IN_VINCULO_SECRETARIA_EDUCACAO        5842 non-null   float64
3   IN_VINCULO_SEGURANCA_PUBLICA           5842 non-null   float64
4   IN_VINCULO_SECRETARIA_SAUDE            5842 non-null   float64
5   IN_VINCULO_OUTRO_ORGAO                 5842 non-null   float64
6   IN_PODER_PUBLICO_PARCERIA              8013 non-null   float64
7   TP_PODER_PUBLICO_PARCERIA              184 non-null    float64
dtypes: float64(6), object(2)
memory usage: 621.1+ KB
```

```
df['IN_VINCULO_SECRETARIA_EDUCACAO'].value_counts()
```

```
IN_VINCULO_SECRETARIA_EDUCACAO
1.0      5839
0.0         3
Name: count, dtype: int64
```

```
# df.groupby('IN_PROF_ADMINISTRATIVOS')['QT_PROF_ADMINISTRATIVOS'].sum()
# df_admin = df[df['IN_PROF_ADMINISTRATIVOS'] == 1]

df_filtro_profissionais = df[
    (df['IN_PROF_ADMINISTRATIVOS'] == 1) &
    (df['IN_PROF_SERVICOS_GERAIS'] == 1) &
    (df['IN_PROF_BIBLIOTECARIO'] == 1) &
    (df['IN_PROF_SAUDE'] == 1) &
    (df['IN_PROF_COORDENADOR'] == 1) &
    (df['IN_PROF_FONAUDIOLOGO'] == 1) &
    (df['IN_PROF_NUTRICIONISTA'] == 1) &
    (df['IN_PROF_PSICOLOGO'] == 1) &
    (df['IN_PROF_ALIMENTACAO'] == 1) &
    (df['IN_PROF_PEDAGOGIA'] == 1) &
    (df['IN_PROF_SECRETARIO'] == 1) &
    (df['IN_PROF_SEGURANCA'] == 1) &
    (df['IN_PROF_MONITORES'] == 1) &
    (df['IN_PROF_ASSIST_SOCIAL'] == 1) &
    (df['IN_PROF_TRAD_LIBRAS'] == 1)
]
```

```
# df.groupby('IN_PROF_ADMINISTRATIVOS')['QT_PROF_ADMINISTRATIVOS'].sum()
# df_admin = df[df['IN_PROF_ADMINISTRATIVOS'] == 1]

df_filtro_profissionais2 = df[
    (df['IN_PROF_ADMINISTRATIVOS'] == 1) |
    (df['IN_PROF_SERVICOS_GERAIS'] == 1) |
    (df['IN_PROF_BIBLIOTECARIO'] == 1) |
    (df['IN_PROF_SAUDE'] == 1) |
    (df['IN_PROF_COORDENADOR'] == 1) |
    (df['IN_PROF_FONAUDIOLOGO'] == 1) |
    (df['IN_PROF_NUTRICIONISTA'] == 1) |
    (df['IN_PROF_PSICOLOGO'] == 1) |
    (df['IN_PROF_ALIMENTACAO'] == 1) |
    (df['IN_PROF_PEDAGOGIA'] == 1) |
    (df['IN_PROF_SECRETARIO'] == 1) |
    (df['IN_PROF_SEGURANCA'] == 1) |
    (df['IN_PROF_MONITORES'] == 1) |
    (df['IN_PROF_ASSIST_SOCIAL'] == 1) |
    (df['IN_PROF_TRAD_LIBRAS'] == 1)
]
```

```
df[['IN_PROF_ADMINISTRATIVOS', 'IN_PROF_SERVICOS_GERAIS', 'IN_PROF_BIBLIOTECARIO']]
```

border="1" class="dataframe">

	IN_PROF_ADMINISTRATIVOS	IN_PROF_SERVICOS_GERAIS	IN_PROF_BIBLIOTECARIO
0	NaN	NaN	NaN
1	1.0	1.0	0.0
2	1.0	1.0	0.0
3	1.0	1.0	0.0
4	NaN	NaN	NaN
...
9931	1.0	1.0	0.0
9932	NaN	NaN	NaN
9933	1.0	1.0	0.0
9934	1.0	1.0	0.0
9935	1.0	1.0	0.0

9936 rows × 3 columns

```
df_filtro_profissionais2
```

[illegible]

9				Pe																	
9	20	No	2	rn	PE	26	Xe	26	Re	26		5.	0.	0.	0.	0.	0.	0.	0.	0.	0.
3	24	rd		a			xé	16	cif	01	...	0	0	0	0	0	0	0	0	0	0
1		este		mbu			u	50	e												
		te		co				6													
9				Pe																	
9	20	No	2	rn	PE	26	Xe	26	Re	26		7.	0.	0.	0.	0.	0.	0.	0.	0.	0.
3	24	rd		a			xé	16	cif	01	...	0	0	0	0	0	0	0	0	0	0
3		este		mbu			u	50	e												
		te		co				6													
9				Pe																	
9	20	No	2	rn	PE	26	Xe	26	Re	26		11	4.	0.	0.	0.	0.	4.	4.	0.	0.
3	24	rd		a			xé	16	cif	010	0	0	0	0	0	0	0	0	0
4		este		mbu			u	50	e												
		te		co				6													
9				Pe																	
9	20	No	2	rn	PE	26	Xe	26	Re	26		12	0.	0.	0.	0.	0.	0.	0.	0.	0.
3	24	rd		a			xé	16	cif	010	0	0	0	0	0	0	0	0	0
5		este		mbu			u	50	e												
		te		co				6													

7752 rows × 426 columns

```
df_filtro_profissionais2[['NO_MUNICIPIO', 'NO_ENTIDADE', 'IN_PROF_PSIKOLOGO']]
```

border="1" class="dataframe">

	NO_MUNICIPIO	NO_ENTIDADE	IN_PROF_PSILOGO
1	Abreu e Lima	EM - CENTRO COMU E EDUC ISAAC MARTINS RODRIGUES	0.0
2	Abreu e Lima	EM - CENTRO COM EDUCACIONAL WILIBALDO DE FRANC...	0.0
3	Abreu e Lima	ESCOLA DE REFERENCIA EM ENSINO FUNDAMENTAL GEN...	0.0
5	Abreu e Lima	ESCOLA MARECHAL COSTA E SILVA	0.0
7	Abreu e Lima	EM - ESCOLA MUNICIPAL DE ENGENHO NOVO	0.0
...
9930	Xexéu	ESCOLA MUNICIPAL TEREZINHA BARBOSA DA SILVA	0.0
9931	Xexéu	ESCOLA MUNICIPAL JOAO FOGUETEIRO	1.0
9933	Xexéu	CRECHE MUNICIPAL PROFESSORA ETIENE LINS BARRETO	1.0
9934	Xexéu	ESCOLA MUNICIPAL PROFESSORA ISABEL DE CASSIA M...	1.0
9935	Xexéu	ESCOLA MUNICIPAL DR JOSE HAMILTON LINS	0.0

7752 rows × 3 columns

```
df.isnull().sum().sort_values(ascending=False)
```

```
CO_LINGUA_INDIGENA_3    9936
CO_ESCOLA_SEDE_VINCULADA 9933
CO_LINGUA_INDIGENA_2    9931
IN_FORMA_CONT_ES_COOP_TEC_FIN 9926
IN_FORMA_CONT_ES_TERM_FOMENTO 9926
...
SG_UF                    0
NO_UF                    0
CO_REGIAO                0
NO_REGIAO                0
NU_ANO_CENSO             0
Length: 426, dtype: int64
```

```
# Ver colunas que têm mais de X% de valores nulos (exemplo: mais de 50%)
porcentagem_nulos = df.isnull().mean() * 100 # Em porcentagem
colunas_com_muitos_nulos = porcentagem_nulos[porcentagem_nulos > 50]
print(colunas_com_muitos_nulos)
```

```
TP_CATEGORIA_ESCOLA_PRIVADA      78.150161
DS_COMPLEMENTO                    77.425523
TP_PODER_PUBLICO_PARCERIA         98.148148
IN_FORMA_CONT_TERMOS_COLABORA     98.148148
IN_FORMA_CONT_TERMOS_FOMENTO      98.148148
IN_FORMA_CONT_ACORDO_COOP         98.148148
IN_FORMA_CONT_PRESTACAO_SERV      98.148148
IN_FORMA_CONT_COOP_TEC_FIN        98.148148
IN_FORMA_CONT_CONSORCIO_PUB       98.148148
IN_FORMA_CONT_MU_TERMOS_COLAB     98.248792
IN_FORMA_CONT_MU_TERMOS_FOMENTO   98.248792
IN_FORMA_CONT_MU_ACORDO_COOP      98.248792
IN_FORMA_CONT_MU_PREST_SERV       98.248792
IN_FORMA_CONT_MU_COOP_TEC_FIN     98.248792
IN_FORMA_CONT_MU_CONSORCIO_PUB    98.248792
IN_FORMA_CONT_ES_TERMOS_COLAB     99.899356
IN_FORMA_CONT_ES_TERMOS_FOMENTO   99.899356
IN_FORMA_CONT_ES_ACORDO_COOP      99.899356
IN_FORMA_CONT_ES_PREST_SERV       99.899356
IN_FORMA_CONT_ES_COOP_TEC_FIN     99.899356
IN_FORMA_CONT_ES_CONSORCIO_PUB    99.899356
IN_MANT_ESCOLA_PRIVADA_EMP        78.150161
IN_MANT_ESCOLA_PRIVADA_ONG        78.150161
IN_MANT_ESCOLA_PRIVADA_OSCIP      78.150161
IN_MANT_ESCOLA_PRIV_ONG_OSCIP     78.150161
IN_MANT_ESCOLA_PRIVADA_SIND       78.150161
IN_MANT_ESCOLA_PRIVADA_SIST_S     78.150161
IN_MANT_ESCOLA_PRIVADA_S_FINS     78.150161
NU_CNPJ_ESCOLA_PRIVADA           78.150161
NU_CNPJ_MANTENEDORA              78.150161
CO_ESCOLA_SEDE_VINCULADA         99.969807
CO_IES_OFERTANTE                 99.828905
TP_OCUPACAO_GALPAO              99.647746
TP_INDIGENA_LINGUA              98.379630
CO_LINGUA_INDIGENA_1            99.879227
CO_LINGUA_INDIGENA_2            99.949678
CO_LINGUA_INDIGENA_3            100.000000
IN_RESERVA_PPI                   96.769324
IN_RESERVA_RENDA                 96.769324
IN_RESERVA_PUBLICA               96.769324
IN_RESERVA_PCD                   96.769324
IN_RESERVA_OUTROS                96.769324
IN_RESERVA_NENHUMA               96.769324
dtype: float64
```

```
df[colunas_com_muitos_nulos.index].info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 9936 entries, 0 to 9935
```

```
Data columns (total 43 columns):
```

#	Column	Non-Null Count	Dtype
0	TP_CATEGORIA_ESCOLA_PRIVADA	2171 non-null	float64
1	DS_COMPLEMENTO	2243 non-null	object
2	TP_PODER_PUBLICO_PARCERIA	184 non-null	float64
3	IN_FORMA_CONT_TERM_COLABORA	184 non-null	float64
4	IN_FORMA_CONT_TERM_FOMENTO	184 non-null	float64
5	IN_FORMA_CONT_ACORDO_COOP	184 non-null	float64
6	IN_FORMA_CONT_PRESTACAO_SERV	184 non-null	float64
7	IN_FORMA_CONT_COOP_TEC_FIN	184 non-null	float64
8	IN_FORMA_CONT_CONSORCIO_PUB	184 non-null	float64
9	IN_FORMA_CONT_MU_TERM_COLAB	174 non-null	float64
10	IN_FORMA_CONT_MU_TERM_FOMENTO	174 non-null	float64
11	IN_FORMA_CONT_MU_ACORDO_COOP	174 non-null	float64
12	IN_FORMA_CONT_MU_PREST_SERV	174 non-null	float64
13	IN_FORMA_CONT_MU_COOP_TEC_FIN	174 non-null	float64
14	IN_FORMA_CONT_MU_CONSORCIO_PUB	174 non-null	float64
15	IN_FORMA_CONT_ES_TERM_COLAB	10 non-null	float64
16	IN_FORMA_CONT_ES_TERM_FOMENTO	10 non-null	float64
17	IN_FORMA_CONT_ES_ACORDO_COOP	10 non-null	float64
18	IN_FORMA_CONT_ES_PREST_SERV	10 non-null	float64
19	IN_FORMA_CONT_ES_COOP_TEC_FIN	10 non-null	float64
20	IN_FORMA_CONT_ES_CONSORCIO_PUB	10 non-null	float64
21	IN_MANT_ESCOLA_PRIVADA_EMP	2171 non-null	float64
22	IN_MANT_ESCOLA_PRIVADA_ONG	2171 non-null	float64
23	IN_MANT_ESCOLA_PRIVADA_OSCIP	2171 non-null	float64
24	IN_MANT_ESCOLA_PRIV_ONG_OSCIP	2171 non-null	float64
25	IN_MANT_ESCOLA_PRIVADA_SIND	2171 non-null	float64
26	IN_MANT_ESCOLA_PRIVADA_SIST_S	2171 non-null	float64
27	IN_MANT_ESCOLA_PRIVADA_S_FINS	2171 non-null	float64
28	NU_CNPJ_ESCOLA_PRIVADA	2171 non-null	float64
29	NU_CNPJ_MANTENEDORA	2171 non-null	float64
30	CO_ESCOLA_SEDE_VINCULADA	3 non-null	float64
31	CO_IES_OFERTANTE	17 non-null	float64
32	TP_OCUPACAO_GALPAO	35 non-null	float64
33	TP_INDIGENA_LINGUA	161 non-null	float64
34	CO_LINGUA_INDIGENA_1	12 non-null	float64
35	CO_LINGUA_INDIGENA_2	5 non-null	float64
36	CO_LINGUA_INDIGENA_3	0 non-null	float64
37	IN_RESERVA_PPI	321 non-null	float64
38	IN_RESERVA_RENDA	321 non-null	float64
39	IN_RESERVA_PUBLICA	321 non-null	float64
40	IN_RESERVA_PCD	321 non-null	float64
41	IN_RESERVA_OUTROS	321 non-null	float64
42	IN_RESERVA_NENHUMA	321 non-null	float64

```
dtypes: float64(42), object(1)
```

```
memory usage: 3.3+ MB
```