

```
# Importando pandas  
import pandas as pd
```

```
# Importando os dados  
df = pd.read_csv("dados/microdados_censo_PE_2024.csv")
```

```
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">

	NU_ANO_CENSO	NOREGIAO	COREGIAO	NOUF	SUF	CUF	OMUNICIPIO	COMUNICIPIO	NOREGIAO_GEOG_INTERM	COREGIAO_GEOG_INTERM	...	QTUR_BASD	QTUR_BASN	QTUR_BASD	QTUR_INFINT	QTUR_INFPRINT	QTUR_FUNDAINT	QTUR_FUNDAFINT	QTUR_MEDINT
9931	2024	No rd este	2	Pe rn a m bu co	PE	26	Xe xé u	2616506	Re cif e	2601	...	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9932	2024	No rd este	2	Pe rn a m bu co	PE	26	Xe xé u	2616506	Re cif e	2601	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9933	2024	No rd este	2	Pe rn a m bu co	PE	26	Xe xé u	2616506	Re cif e	2601	...	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9934	2024	No rd este	2	Pe rn a m bu co	PE	26	Xe xé u	2616506	Re cif e	2601	...	11.0	4.0	0.0	0.0	0.0	4.0	4.0	0.0
9935	2024	No rd este	2	Pe rn a m bu co	PE	26	Xe xé u	2616506	Re cif e	2601	...	12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows × 426 columns

```
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.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
```

Se você quiser substituir por 0 somente em colunas numéricas, e deixar outras colunas (como object, string, datetime, etc.) intactas.

```
df[df.select_dtypes(include='number').columns] =  
df.select_dtypes(include='number').fillna(0)
```

TRATANDO OS DADOS NULOS

```
df.fillna(0, inplace=True)
```

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

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

Algumas perguntas para a análise:

- Quais escolas tem biblioteca, quadra, banheiros, area verde (ou uma área de lazer)?
- Quais escolas tem laboratórios (de química, de informática, etc)?
- Quais os meios mais comuns de água nas escolas?
- E para energia?

```
# Contagem de dependencia administrativas das escolas (Se são públicas ou privadas)
df['TP_DEPENDENCIA'].value_counts()
```

```
TP_DEPENDENCIA
3      5720
4      3101
2      1089
1         26
Name: count, dtype: int64
```

Diferença entre as escolas:

- A escola municipal faz parte do município, e é ele quem deve responder por ela;
- Já a escola estadual é dever do próprio Estado cuidar dela, com investimento nas refeições, estruturas físicas, concurso para professores, etc.
- Por sua vez, uma escola federal é de responsabilidade do país, por isso vemos os "Institutos Federais de Educação" os IFs.
- E por fim, as escolas privadas não recebem dinheiro público, mas sim dos donos da escola, da mensalidade dos alunos, etc.

1 - Federal | 2 - Estadual | 3 - Municipal | 4 - Privada

```
df_copy = df.copy()

df_copy['CATEGORIA_ESCOLA'] = df_copy['TP_DEPENDENCIA'].map({
    1: 'Federal',
    2: 'Estadual',
    3: 'Municipal',
    4: 'Privada'
})
```

df_copy

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 R A F I C O I N T E R M	C O R E G I A O G E O G R A F I C O I N T E R M	...	Q T U R B A S E S	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 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	C A T E G O R I A E S C O L A
0	2024	Nordeste	2	Pernambuco	PE	26	Abreu e Lima	2600054	Recife	2601	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Privada
1	2024	Nordeste	2	Pernambuco	PE	26	Abreu e Lima	2600054	Recife	2601	...	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Municipal
2	2024	Nordeste	2	Pernambuco	PE	26	Abreu e Lima	2600054	Recife	2601	...	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Municipal
3	2024	Nordeste	2	Pernambuco	PE	26	Abreu e Lima	2600054	Recife	2601	...	9.0	0.0	0.0	0.0	0.0	3.0	0.0	3.0	Estadual
4	2024	Nordeste	2	Pernambuco	PE	26	Abreu e Lima	2600054	Recife	2601	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Privada
...
9931	2024	Nordeste	2	Pernambuco	PE	26	Xexéu	2616506	Recife	2601	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Municipal

9932	2024	No rd este	2	Pe rn a m bu co	PE	26	Xe xé u	2616506	Re cif e	2601	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Pri va da
9933	2024	No rd este	2	Pe rn a m bu co	PE	26	Xe xé u	2616506	Re cif e	2601	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	M un ici pa l
9934	2024	No rd este	2	Pe rn a m bu co	PE	26	Xe xé u	2616506	Re cif e	2601	...	4.0	0.0	0.0	0.0	0.0	4.0	4.0	0.0	M un ici pa l
9935	2024	No rd este	2	Pe rn a m bu co	PE	26	Xe xé u	2616506	Re cif e	2601	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	M un ici pa l

9936 rows × 427 columns

```
df_publicas = df_copy[df_copy['CATEGORIA_ESCOLA'].isin(['Federal', 'Estadual', 'Municipal'])]
df_privadas = df_copy[df_copy['CATEGORIA_ESCOLA'] == 'Privada']
```

```
df_publicas
```


[illegible]

6835 rows x 427 columns

```
df_privadas
```

[illegible]

9 9 0 7	20 24	No rd este	2	Pe rn a m bu co	PE	26	Vit óri a de Sa nt o An tã o	26 16 40 7	Re cif e	26 01	...	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	Pri va da
9 9 0 8	20 24	No rd este	2	Pe rn a m bu co	PE	26	Vit óri a de Sa nt o An tã o	26 16 40 7	Re cif e	26 01	...	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	Pri va da
9 9 0 9	20 24	No rd este	2	Pe rn a m bu co	PE	26	Vit óri a de Sa nt o An tã o	26 16 40 7	Re cif e	26 01	...	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	Pri va da
9 9 3 2	20 24	No rd este	2	Pe rn a m bu co	PE	26	Xe xé u	26 16 50 6	Re cif e	26 01	...	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0	Pri va da

3101 rows × 427 columns

Estrutura das escolas

Quais escolas tem bibliotecas (e/ ou sala de leituras)?

Em escolas públicas (Federais, Municipais e Estaduais)

```
df_publicas['IN_BIBLIOTECA'].value_counts()
```

```
IN_BIBLIOTECA
0.0      4567
```

```
1.0    2268
Name: count, dtype: int64
```

```
df_publicas['IN_BIBLIOTECA_SALA_LEITURA'].value_counts()
```

```
IN_BIBLIOTECA_SALA_LEITURA
0.0    3800
1.0    3035
Name: count, dtype: int64
```

```
# escolas_com_biblioteca = df[df['IN_BIBLIOTECA'] == 1]
```

```
df_bibli_pub = df_publicas[
    (df_publicas['IN_BIBLIOTECA'] == 1) &
    (df_publicas['IN_BIBLIOTECA_SALA_LEITURA'] == 1)
]
```

```
df_resumo_bibli_pub = df_bibli_pub.groupby('NO_MUNICIPIO')['IN_BIBLIOTECA'].count()
df_resumo_bibli_pub = df_resumo_bibli_pub.sort_values(ascending=False)
```

```
# df_resumo_bibli[df_resumo_bibli.index == 'Vicência']
```

```
df_resumo_bibli_pub.head() # Municípios com maior numeros de escolas com bibliotecas
```

```
NO_MUNICIPIO
Recife          332
Jaboatão dos Guararapes  105
Caruaru         79
Petrolina       58
Olinda          56
Name: IN_BIBLIOTECA, dtype: int64
```

```
df_resumo_bibli_pub.tail() # Municípios com menor numeros de escolas com bibliotecas
```

```
NO_MUNICIPIO
Salgadinho      1
Solidão         1
Sanharó         1
Terra Nova      1
Terezinha       1
Name: IN_BIBLIOTECA, dtype: int64
```

Em escolas privadas

```
df_privadas['IN_BIBLIOTECA'].value_counts()
```

```
IN_BIBLIOTECA
0.0    1571
1.0    1530
Name: count, dtype: int64
```

```
df_bibli_priv = df_publicas[
    (df_publicas['IN_BIBLIOTECA'] == 1) &
    (df_publicas['IN_BIBLIOTECA_SALA_LEITURA'] == 1)
]

df_resumo_bibli_priv = df_bibli_priv.groupby('NO_MUNICIPIO')
['IN_BIBLIOTECA'].count()
df_resumo_bibli_priv = df_resumo_bibli_priv.sort_values(ascending=False)
```

```
# df_resumo_bibli.to_excel('resumo_bibli.xlsx')
```

```
df2 = df[
    (df['IN_BIBLIOTECA'] == 1) &
    (df['IN_BIBLIOTECA_SALA_LEITURA'] == 1) &
    (df['TP_LOCALIZACAO'] == 2) # 1 - Urbana | 2 = Rural
]

df2_resumo = df2.groupby('NO_MUNICIPIO')['IN_BIBLIOTECA'].count()
df2_resumo = df2_resumo.sort_values(ascending=False)
```

df2_resumo

NO_MUNICIPIO	
Caruaru	22
Carnaubeira da Penha	19
Cabo de Santo Agostinho	17
Petrolina	17
Paudalho	14
..	
Santa Terezinha	1
Tupanatinga	1
Tacaimbó	1
São Vicente Férrer	1
Venturosa	1

Name: IN_BIBLIOTECA, Length: 137, dtype: int64

ESCOLAS COM BIBLIOTECAS E SEM

```
df_com_biblio = df[df['IN_BIBLIOTECA'] == 1] # COM
df_sem_biblio = df[df['IN_BIBLIOTECA'] == 0] # SEM
```

```
# Agrupar e contar por município (ou outra categoria)
```

```
resumo_com = df_com_biblio.groupby('NO_MUNICIPIO').size().rename('COM_BIBLIOTECA')
resumo_sem = df_sem_biblio.groupby('NO_MUNICIPIO').size().rename('SEM_BIBLIOTECA')
```

```
# Juntando os dois dataframes
```

```
df_resumo_bibli_2 = pd.concat([resumo_com, resumo_sem],
axis=1).fillna(0).astype(int)
```

```
df_resumo_bibli_2['TOTAL'] = df_resumo_bibli_2['COM_BIBLIOTECA'] +
df_resumo_bibli_2['SEM_BIBLIOTECA']
```

```
df_resumo_bibli_2 = df_resumo_bibli_2.sort_values('TOTAL', ascending=False)
```

```
df_resumo_bibli_2
```

```
border="1" class="dataframe">
```

NO_MUNICIPIO	COM_BIBLIOTECA	SEM_BIBLIOTECA	TOTAL
Recife	696	539	1235
Jaboatão dos Guararapes	238	276	514
Petrolina	108	271	379
Olinda	170	162	332
Caruaru	147	151	298
...
Ingazeira	1	9	10
Terra Nova	1	8	9
Granito	6	3	9
Solidão	1	5	6
Fernando de Noronha	2	0	2

185 rows × 3 columns

```
# Visualizar a proporção do dataframe
```

```
df_resumo_bibli_2['%_COM_BIBLIOTECA'] = (df_resumo_bibli_2['COM_BIBLIOTECA'] /  
df_resumo_bibli_2['TOTAL']) * 100
```

```
df_resumo_bibli_2['%_SEM_BIBLIOTECA'] = (df_resumo_bibli_2['SEM_BIBLIOTECA'] /  
df_resumo_bibli_2['TOTAL']) * 100
```

```
df_resumo_bibli_2
```

border="1" class="dataframe">

NO_MUNICIPIO	COM_BIBLIOTECA	SEM_BIBLIOTECA	TOTAL	%_COM_BIBLIOTECA	%_SEM_BIBLIOTECA
Recife	696	539	1235	56.356275	43.643725
Jaboatão dos Guararapes	238	276	514	46.303502	53.696498
Petrolina	108	271	379	28.496042	71.503958
Olinda	170	162	332	51.204819	48.795181
Caruaru	147	151	298	49.328859	50.671141
...
Ingazeira	1	9	10	10.000000	90.000000
Terra Nova	1	8	9	11.111111	88.888889
Granito	6	3	9	66.666667	33.333333
Solidão	1	5	6	16.666667	83.333333
Fernando de Noronha	2	0	2	100.000000	0.000000

185 rows × 5 columns

Infraestrutura ligada ao esgoto das escolas

Algumas escolas tem mais de uma opção do tipo de esgoto. Devido a isso, fiz duas análises:

1º: Mostra e apresenta a abordagem resumida (sem duplicidade)

2º: Apresente a realidade detalhada em separado

```
colunas_esgoto = [  
    'TP_LOCALIZACAO',  
    'IN_ESGOTO_REDE_PUBLICA',  
    'IN_ESGOTO_FOSSA_SEPTICA',  
    'IN_ESGOTO_FOSSA_COMUM',  
    'IN_ESGOTO_FOSSA',  
    'IN_ESGOTO_INEXISTENTE'  
]
```

1º Primeira Análise

```
# Contar quantas escolas têm cada infraestrutura (valor == 1)  
df_resumo_esgoto = df.groupby('NO_MUNICIPIO')[colunas_esgoto[1:]].sum()
```

```
df_resumo_esgoto['TOTAL'] = df_resumo_esgoto[colunas_esgoto[1]] +
df_resumo_esgoto[colunas_esgoto[2]] + df_resumo_esgoto[colunas_esgoto[3]] +
df_resumo_esgoto[colunas_esgoto[4]] + df_resumo_esgoto[colunas_esgoto[5]]

df_resumo_esgoto = df_resumo_esgoto.sort_values('TOTAL', ascending=False)
```

```
df_resumo_esgoto
```

```
border="1" class="dataframe">
```

NO_MUNICIPPIO	IN_ESGOTO_REDE_PUBLICA	IN_ESGOTO_FOSSA_SEPTICA	IN_ESGOTO_FOSSA_COMUM	IN_ESGOTO_FOSSA	IN_ESGOTO_INEXISTENTE	TOTAL
Recife	836.0	150.0	105.0	255.0	0.0	1346.0
Jaboatão dos Guararapes	243.0	70.0	75.0	145.0	2.0	535.0
Petrolina	223.0	104.0	19.0	123.0	1.0	470.0
Caruaru	183.0	68.0	11.0	79.0	0.0	341.0
Olinda	211.0	20.0	23.0	43.0	4.0	301.0
...
Solidão	3.0	2.0	2.0	4.0	0.0	11.0
Tuparetama	6.0	0.0	2.0	2.0	0.0	10.0
Granito	6.0	1.0	1.0	2.0	0.0	10.0
Brejinho	2.0	1.0	2.0	3.0	0.0	8.0
Fernando de Noronha	2.0	1.0	0.0	1.0	0.0	4.0

185 rows × 6 columns

```
# df[colunas_esgoto]['IN_ESGOTO_INEXISTENTE'].value_counts()
df[colunas_esgoto]
```


border="1" class="dataframe">

	IN_ESGOTO_	IN_ESGOTO_	IN_ESGOTO_	IN_ESGOTO_	IN_ESGOTO_I
	RÉDE_PUBLI	FOSSA_SEPTI	FOSSA_COM	FOSSA	NEXISTENTE
	CA	CA	UM		
0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	1.0	1.0	0.0
2	0.0	0.0	1.0	1.0	0.0
3	0.0	0.0	1.0	1.0	0.0
4	0.0	0.0	0.0	0.0	0.0
...
9931	1.0	0.0	0.0	0.0	0.0
9932	0.0	0.0	0.0	0.0	0.0
9933	1.0	0.0	0.0	0.0	0.0
9934	1.0	0.0	0.0	0.0	0.0
9935	1.0	0.0	0.0	0.0	0.0

9936 rows × 5 columns

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

```
TP_LOCALIZACAO
1      6169
2      3767
Name: count, dtype: int64
```