ANÁLISE DA BASE DO BRASILEIRÃO

Importar biblioteca

```
import pandas as pd
```

Leitura de Arquivos

```
[2] full = pd.read_csv('campeonato-brasileiro-full.csv')
   gols = pd.read_csv('campeonato-brasileiro-gols.csv')
   estatisticas = pd.read_csv('campeonato-brasileiro-estatisticas-full.csv')
   cartoes = pd.read_csv('campeonato-brasileiro-cartoes.csv')
```

Tratamento dos dados (filtro)

```
full_2023 = full[full['data'].str.contains('2023')]

[4] # Ensure ID columns in all DataFrames are of the same type (int64):
    full_2023['ID'] = full_2023['ID'].astype('int64') # Already int64, but for consistency
    gols['partida_id'] = gols['partida_id'].astype('int64')
    estatisticas['partida_id'] = estatisticas['partida_id'].astype('int64')
    cartoes['partida_id'] = cartoes['partida_id'].astype('int64')

# Perform the merge using left_on and right_on:
    gols_2023 = pd.merge(full_2023, gols, left_on='ID', right_on='partida_id', how='inner')
    estatisticas_2023 = pd.merge(full_2023, estatisticas, left_on='ID', right_on='partida_id', how='inner')
    cartoes_2023 = pd.merge(full_2023, cartoes, left_on='ID', right_on='partida_id', how='inner')
```

Análise descritiva

Head() and tail()

full_2023

Đ	ID	rodata	data	hora	mandante	visitante	formacao_mandante	formacao_visitante	tecnico_mandante	tecnico_visitante	vencedor	arena	mandante_Placar	visitante_Placar	mandante_Estado	visitante_Estado
802	8026	1	15/04/2023	16:01	Palmeiras	Cuiaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira	Palmeiras	Allianz Parque	2	1	SP	MT
802	8027	1	15/04/2023	16:02	America- MG	Fluminense	4-2-3-1	4-2-2-2	V. Carmo Mancini	F. Diniz Silva	Fluminense	Estádio Raimundo Sampaio	0	3	MG	RJ
802	8029	1	15/04/2023	18:31	Bragantino	Bahia	4-2-3-1	3-4-3	P. Faria Caixinha	R. Alves Paiva	Bragantino	Nabizão	2	1	SP	BA
802	8030	1	15/04/2023	18:31	Athletico- PR	Goias	4-2-3-1	4-2-3-1	P. Turra	E. Rodrigues de Ávila	Athletico- PR	Ligga Arena	2	0	PR	GO
802	8031	1	15/04/2023	18:32	Fortaleza	Internacional	4-2-2-2	4-2-3-1	J. Vojvoda	L. Venker de Menezes		Arena Castelão	1	1	CE	RS
840	8401	38	06/12/2023	21:32	Bahia	Atletico-MG	3-4-2-1	4-4-2	R. Ceni	L. Scolari	Bahia	Itaipava Arena Fonte Nova	4	1	BA	MG
840	8402	38	06/12/2023	21:32	Cuiaba	Athletico-PR	5-3-2	3-4-3	A. Cardoso de Oliveira	W. Carvalho	Cuiaba	Arena Pantanal	3	0	MT	PR
840:	8403	38	06/12/2023	21:32	Santos	Fortaleza	3-4-1-2	4-2-3-1	M. Fernandes	J. Vojvoda	Fortaleza	Estádio Urbano Caldeira	1	2	SP	CE
840	8404	38	06/12/2023	21:32	Sao Paulo	Flamengo	4-2-3-1	4-2-3-1	D. Silvestre Júnior	A. Bacchi	Sao Paulo	Morumbi	1	0	SP	RJ
840	8405	38	06/12/2023	21:32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Caixinha	Vasco	Estádio São Januário	2	1	RJ	SP
380 r	ows × 16	columns														

gols_2023

	ID rodata_x	data	hora	mandante	visitante	formacao_mandante	formacao_visitante	tecnico_mandante	tecnico_visitante	 mandante_Placar	visitante_Placar	mandante_Estad	visitante_Estad	o partida_id	rodata_y	clube	atleta	minuto	tipo_de_gol
0 80	26 1	15/04/2023	16:01	Palmeiras	Cuiaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira	2	1	S	. м	T 8026	1	Cuiaba	Raniele	45+7	NaN
1 80	26 1	15/04/2023	16:01	Palmeiras	Cuiaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira	2	1	S	Р М	T 8026	1	Palmeira	Endrick	5	NaN
2 80	26 1	15/04/2023	16:01	Palmeiras	Cuiaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira	2	1	S	. м	T 8026	1	Palmeira	José Manuel Alberto López	65	NaN
3 80	27 1	15/04/2023	16:02	America- MG	Fluminense	4-2-3-1	4-2-2-2	V. Carmo Mancini	F. Diniz Silva	0	3	M	3 R	J 8027	1	Fluminense	Germán Cano	52	NaN
4 80	27 1	15/04/2023	16:02	America- MG	Fluminense	4-2-3-1	4-2-2-2	V. Carmo Mancini	F. Diniz Silva	0	3	M	В В	J 8027	1	Fluminense	John Kennedy	60	NaN
941 84	03 38	06/12/2023	21:32	Santos	Fortaleza	3-4-1-2	4-2-3-1	M. Fernandes	J. Vojvoda	1	2	S	· c	E 8403	38	Santo	Messias Rodrígues da Silva Júnior	58	NaN
942 84	04 38	06/12/2023	21:32	Sao Paulo	Flamengo	4-2-3-1	4-2-3-1	D. Silvestre Júnior	A. Bacchi	1	0	S	, н	J 8404	38	Sao Paulo	Luciano da Rocha Neves	26	NaN
943 84	05 38	06/12/2023	21:32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Caixinha	2	1	R	J S	P 8405	38	Vasco	Sergio Antonio De Luiz Junior	82	NaN
944 84	05 38	06/12/2023	21:32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Caixinha	2	1	R	J S	P 8405	38	Bragantino	Léo Ortiz	63	NaN
945 84	05 38	06/12/2023	21:32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Caixinha	2	1	R	J S	P 8405	38	Vasco	Paulinho	29	NaN
946 rows	× 22 columns																		

estatisticas_2023

1	ID rodata_x	da	a ho	ra m	nandante	visitante	formacao_mandante	formacao_visitante	tecnico_mandante	tecnico_visitante	с	hutes	chutes_no_alvo	posse_de_bola	passes	precisao_passes	faltas	cartao_amarelo	cartao_vermelho	impedimentos	escanteios
0 802	26 1	15/04/20	3 16:	01 P	Palmeiras	Cuiaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira		12	4	48%	427	79%	12	3	1	2	4
1 802	26 1	15/04/20	3 16:	01 P	Palmeiras	Cuiaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira		17	5	52%	455	84%	23	2	0	3	4
2 802	27 1	15/04/20	3 16:	02 Ame	erica-MG I	Fluminense	4-2-3-1	4-2-2-2	V. Carmo Mancini	F. Diniz Silva		15	8	55%	440	83%	9	2	0	3	
3 802	27 1	15/04/20	3 16:	02 Ame	erica-MG I	Fluminense	4-2-3-1	4-2-2-2	V. Carmo Mancini	F. Diniz Silva		17	2	45%	333	79%	25	3	0	2	1
4 802	19 1	15/04/20	3 18:	31 Br	ragantino	Bahia	4-2-3-1	3-4-3	P. Faria Caixinha	R. Alves Paiva		9	3	43%	343	72%	13	1	0	3	
755 840	38	06/12/20	3 21:	32	Santos	Fortaleza	3-4-1-2	4-2-3-1	M. Fernandes	J. Vojvoda		15	6	55%	353	77%	15	2	0	4	1
756 840	14 38	06/12/20	3 21:	32 S	ao Paulo	Flamengo	4-2-3-1	4-2-3-1	D. Silvestre Júnior	A. Bacchi		9	3	42%	406	84%	13	0	0	2	
757 840	14 38	06/12/20	3 21:	32 S	ao Paulo	Flamengo	4-2-3-1	4-2-3-1	D. Silvestre Júnior	A. Bacchi		10	2	58%	535	89%	16	3	0	0	9
758 840	5 38	06/12/20	3 21:	32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Caixinha		13	6	45%	321	78%	8	3	0	0	
759 840)5 38	06/12/20	3 21:	32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Caixinha		9	2	55%	398	80%	14	1	1	1	:
60 rows x	29 columns																				

cartoes_2023

	ID rodata_x	data	hora	mandante	visitante	formacao_mandante	formacao_visitante	tecnico_mandante	tecnico_visitante	 mandante_Estado	visitante_Estado	o partida	_id re	data_y	clube	cartao	atlet	a num_ca	misa	posicao	minuto
0 80	26 1	15/04/2023	16:01	Palmeiras	Cuiaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira	SP	MT	т 8	026	1	Cuiaba	Vermelho	Filipe Augusto Carvalho Souz	а	5.0 Me	eio-campo	57
1 80	26 1	15/04/2023	16:01	Palmeiras	Culaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira	SP	MT	т 8	026	1	Palmeiras	Amarelo	José Manuel Alberto Lópe	z	18.0 Me	elo-campo	37
2 80	26 1	15/04/2023	16:01	Palmeiras	Cuiaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira	SP	M	т 8	026	1	Cuiaba	Amarelo	Mateusinh	0	98.0	Zagueiro	41
3 80	26 1	15/04/2023	16:01	Palmeiras	Cuiaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira	SP	MT	т 8	026	1	Cuiaba	Amarelo	Deyverso	n	16.0	Atacante	
4 80	26 1	15/04/2023	16:01	Palmeiras	Culaba	4-2-3-1	4-2-3-1	A. Moreira Ferreira	I. Abreu Vieira	SP	M	т 8	026	1	Palmeiras	Amarelo	Jhonatan dos Santos Ros	a	40.0 Me	eio-campo	66
2113 84	05 38	06/12/2023	21:32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Catrinha	RJ	SF	8	405	38	Vasco	Amarelo	Gabriel Pe	c	11.0	Atacante	90+5
2114 84	05 38	06/12/2023	21:32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Caldinha	RJ	SF	8	405	38	Vasco	Amarelo	Paulinh	0	18.0 Me	eio-campo	51
2115 84	05 38	06/12/2023	21:32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Calxinha	RJ	SF	· 8	405	38	Vasco	Amarelo	Gary Mede	ol .	17.0	Zagueiro	84
2116 84	05 38	06/12/2023	21:32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Caixinha	RJ	SF	8	405	38	Bragantino	Amarelo	Eduardo Sash	a	19.0	Atacante	17
2117 84	05 38	06/12/2023	21:32	Vasco	Bragantino	4-4-2	4-3-3	R. Díaz	P. Faria Catrinha	RJ	SF	8	405	38	Bragantino	Vermelho	Léo Realp	e	2.0	Zagueiro	75
2118 rows	× 24 columns																				

.info()

full_2023

Informações da tabela 'full_2023':
<class 'pandas.core.frame.DataFrame'>
Index: 380 entries, 8025 to 8404
Data columns (total 16 columns):

#	Column	Non-Null Count	Dtype
0	ID	380 non-null	int64
1	rodata	380 non-null	int64
2	data	380 non-null	object
3	hora	380 non-null	object
4	mandante	380 non-null	object
5	visitante	380 non-null	object
6	formacao_mandante	380 non-null	object
7	formacao_visitante	380 non-null	object
8	tecnico_mandante	380 non-null	object
9	tecnico_visitante	380 non-null	object
10	vencedor	380 non-null	object
11	arena	380 non-null	object
12	mandante_Placar	380 non-null	int64
13	visitante_Placar	380 non-null	int64
14	mandante_Estado	380 non-null	object
15	visitante_Estado	380 non-null	object

dtypes: int64(4), object(12)
memory usage: 58.6+ KB

None

gols 2023

Informações da tabela 'gols_2023':
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 946 entries, 0 to 945
Data columns (total 22 columns):

Non-Null Count Dtype # Column --------0 ID 946 non-null int64 946 non-null int64 946 non-null object 1 rodata_x object 2 data 946 non-null 3 hora object 946 non-null object 4 mandante 946 non-null object 5 visitante 6 formacao_mandante 946 non-null object 7 formacao_visitante 946 non-null object 8 tecnico_mandante 946 non-null object 9 tecnico_visitante 946 non-null object 10 vencedor 946 non-null object 10 vencedor 946 non-null object 11 arena 12 mandante_Placar 946 non-null int64 13 visitante_Placar 946 non-null int64
14 mandante_Estado 946 non-null object
15 visitante_Estado 946 non-null object 16 partida_id 946 non-null int64 17 rodata_y 946 non-null int64 17 rodata_y 18 clube 946 non-null object 19 atleta 946 non-null object 20 minuto 946 non-null object 21 tipo_de_gol 118 non-null object

dtypes: int64(6), object(16) memory usage: 162.7+ KB

None

estatisticas 2023

#	Column	Non-Null Count	Dtype
0	ID	760 non-null	int64
1	rodata_x	760 non-null	int64
2	data	760 non-null	object
3	hora	760 non-null	object
4	mandante	760 non-null	object
5	visitante	760 non-null	object
6	formacao_mandante	760 non-null	object
7	formacao_visitante	760 non-null	object
8	tecnico_mandante	760 non-null	object
9	tecnico_visitante	760 non-null	object
10	vencedor	760 non-null	object
11	arena	760 non-null	object
12	mandante_Placar	760 non-null	int64
13	visitante_Placar	760 non-null	int64
14	mandante_Estado	760 non-null	object
15	visitante_Estado	760 non-null	object
16	partida_id	760 non-null	int64
17	rodata_y	760 non-null	int64
18	clube	760 non-null	object
19	chutes	760 non-null	int64
20	chutes_no_alvo	760 non-null	int64
21	posse_de_bola	760 non-null	object
22	passes	760 non-null	int64
23	precisao_passes	760 non-null	object
24	faltas	760 non-null	int64
25	cartao_amarelo	760 non-null	int64
26	cartao_vermelho	760 non-null	int64
27	impedimentos	760 non-null	int64
28	escanteios	760 non-null	int64
dtvp	es: int64(14), obiec	t(15)	

dtypes: int64(14), object(15) memory usage: 172.3+ KB

None

cartoes 2023

```
Informações da tabela 'cartoes_2023':
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2118 entries, 0 to 2117
Data columns (total 24 columns):
# Column
                       Non-Null Count Dtype
                       -----
0 ID
                      2118 non-null
                                      int64
                       2118 non-null
1
    rodata_x
                       2118 non-null
                                      object
2
    data
                       2118 non-null
4
   mandante
                      2118 non-null
                                      object
    visitante
                       2118 non-null
                                       object
   formacao_mandante 2118 non-null
 6
                                       object
   formacao_visitante 2118 non-null
8 tecnico_mandante 2118 non-null
9 tecnico_visitante 2118 non-null
                                       object
10 vencedor
                       2118 non-null
                                       object
11 arena
                       2118 non-null
                     2118 non-null
2118 non-null
12 mandante_Placar
 13 visitante_Placar
14 mandante_Estado
                       2118 non-null
                                      object
 15 visitante_Estado
                      2118 non-null
16 partida_id
                       2118 non-null
 17 rodata_y
                        2118 non-null
18 clube
                       2118 non-null
                                       object
 19 cartao
                      2118 non-null object
 20 atleta
                       2118 non-null
                                       object
 21 num_camisa
                       2118 non-null
                                       float64
22 posicao
                      2118 non-null
                                      object
23 minuto
                        2118 non-null
dtypes: float64(1), int64(6), object(17)
memory usage: 397.2+ KB
None
```

Análise quantitativa

.describe()

full 2023

full_2	023.describe()		
	ID	rodata	mandante_Placar	visitante_Placar
count	380.000000	380.000000	380.000000	380.000000
mean	8215.500000	19.500000	1.418421	1.071053
std	109.840794	10.980313	1.189744	1.090958
min	8026.000000	1.000000	0.000000	0.000000
25%	8120.750000	10.000000	1.000000	0.000000
50%	8215.500000	19.500000	1.000000	1.000000
75%	8310.250000	29.000000	2.000000	2.000000
max	8405.000000	38.000000	7.000000	6.000000

gols_2023.describe()

	ID	rodata_x	mandante_Placar	visitante_Placar	partida_id	rodata_y	minuto
count	946.000000	946.000000	946.000000	946.000000	946.000000	946.000000	816.000000
mean	8215.052854	19.450317	1.997886	1.560254	8215.052854	19.450317	47.127451
std	111.469774	11.149687	1.325371	1.272861	111.469774	11.149687	25.602405
min	8026.000000	1.000000	0.000000	0.000000	8026.000000	1.000000	1.000000
25%	8117.000000	10.000000	1.000000	1.000000	8117.000000	10.000000	26.000000
50%	8219.000000	20.000000	2.000000	1.000000	8219.000000	20.000000	48.000000
75%	8309.000000	29.000000	3.000000	2.000000	8309.000000	29.000000	69.000000
max	8405.000000	38.000000	7.000000	6.000000	8405.000000	38.000000	90.000000

estatisticas_2023

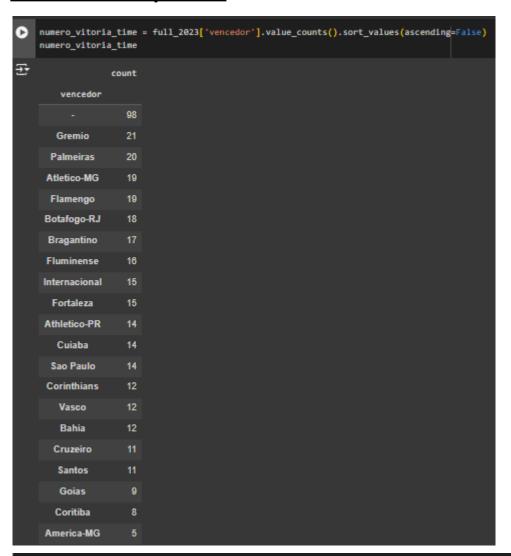
estatis	ticas_2023.d	escribe()												
	ID	rodata_x	mandante_Placar	visitante_Placar	partida_id	rodata_y	chutes	chutes_no_alvo	passes	faltas	cartao_amarelo	cartao_vermelho	impedimentos	escanteios
count	760.000000	760.000000	760.000000	760.000000	760.000000	760.000000	760.000000	760.000000	760.000000	760.000000	760.000000	760.000000	760.000000	760.000000
mean	8215.500000	19.500000	1.418421	1.071053	8215.500000	19.500000	13.514474	4.555263	428.434211	14.336842	2.588158	0.142105	1.746053	5.285526
std	109.768411	10.973078	1.188960	1.090239	109.768411	10.973078	5.354107	2.363446	105.463221	4.295687	1.450134	0.385256	1.560374	2.973542
min	8026.000000	1.000000	0.000000	0.000000	8026.000000	1.000000	2.000000	0.000000	195.000000	2.000000	0.000000	0.000000	0.000000	0.000000
25%	8120.750000	10.000000	1.000000	0.000000	8120.750000	10.000000	9.000000	3.000000	353.000000	12.000000	2.000000	0.000000	1.000000	3.000000
50%	8215.500000	19.500000	1.000000	1.000000	8215.500000	19.500000	13.000000	4.000000	420.500000	14.000000	2.000000	0.000000	1.000000	5.000000
75%	8310.250000	29.000000	2.000000	2.000000	8310.250000	29.000000	17.000000	6.000000	505.000000	17.000000	3.000000	0.000000	3.000000	7.000000
max	8405.000000	38.000000	7.000000	6.000000	8405.000000	38.000000	36.000000	18.000000	744.000000	29.000000	8.000000	2.000000	9.000000	17.000000

cartoes_2023

cartoes_2023.describe()

	ID	rodata_x	mandante_Placar	visitante_Placar	partida_id	rodata_y	num_camisa
count	2118.000000	2118.000000	2118.000000	2118.000000	2118.000000	2118.000000	2118.000000
mean	8209.277148	18.874882	1.399906	1.089707	8209.277148	18.874882	21.553352
std	108.167382	10.795559	1.153393	1.072910	108.167382	10.795559	19.155820
min	8026.000000	1.000000	0.000000	0.000000	8026.000000	1.000000	1.000000
25%	8115.250000	9.250000	1.000000	0.000000	8115.250000	9.250000	8.000000
50%	8210.000000	19.000000	1.000000	1.000000	8210.000000	19.000000	17.000000
75%	8298.000000	28.000000	2.000000	2.000000	8298.000000	28.000000	28.000000
max	8405.000000	38.000000	7.000000	6.000000	8405.000000	38.000000	99.000000

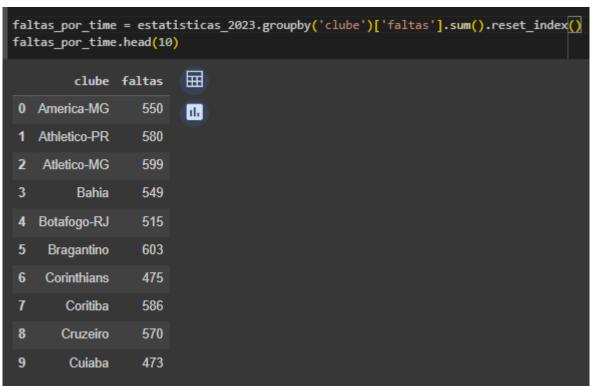
Análise retrospectiva



```
jogos_arena = full_2023['arena'].value_counts().sort_values(ascending=False)
jogos_arena.head(10)
                            count
                     arena
         Maracanã
                               36
Estádio Raimundo Sampaio
                               23
         Mineirão
                               21
  Estádio Urbano Caldeira
                               20
 Itaipava Arena Fonte Nova
                               19
                               19
         Nabizão
       Ligga Arena
                               19
   Estádio Nilton Santos
                               19
    Neo Química Arena
                               19
      Arena Pantanal
                               19
```

```
gols_por_minuto = gols_2023['minuto'].value_counts().sort_index()
gols_por_minuto
        count
minuto
  10
            8
  11
  12
  13
 90+5
        14
 90+6
 90+7
          6
            4
 90+8
 90+9
111 rows × 1 columns
```

```
chutes_no_alvo_por_time = estatisticas_2023.groupby('clube')['chutes_no_alvo'].sum()
chutes_no_alvo_por_time.head(10)
              chutes_no_alvo
       clube
America-MG
                         165
Athletico-PR
                         174
 Atletico-MG
                         173
   Bahia
                         171
Botafogo-RJ
                         179
 Bragantino
                         211
 Corinthians
                         164
  Coritiba
                         152
  Cruzeiro
                         167
   Cuiaba
                         139
```



```
cartoes_por_time = estatisticas_2023.groupby('clube')[['cartao_amarelo', 'cartao_vermelho']].sum().reset_index()
cartoes_por_time.head(10)
                                                     ▦
         clube cartao_amarelo cartao_vermelho
0 America-MG
                             83
                                                     ılı
   Athletico-PR
                            107
    Atletico-MG
 2
                            110
 3
                                                6
         Bahia
                             93
   Botafogo-RJ
                             91
 4
 5
     Bragantino
                            107
 6
    Corinthians
                             92
        Coritiba
 7
                            117
       Cruzeiro
 8
                             97
                                                4
 9
        Cuiaba
                                                3
                             86
```

```
cartoes_por_posicao = cartoes_2023['posicao'].value_counts()
cartoes_por_posicao

count

posicao

Meio-campo 862

Zagueiro 799

Atacante 372

Goleiro 85
```

```
cartoes_primeiro_tempo = cartoes_2023[cartoes_2023['minuto'] <= 45]['cartao'].count()
cartoes_segundo_tempo = cartoes_2023[cartoes_2023['minuto'] > 45]['cartao'].count()

print(f"Cartões no primeiro tempo: {cartoes_primeiro_tempo}")

print(f"Cartões no segundo tempo: {cartoes_segundo_tempo}")

Cartões no primeiro tempo: 628
Cartões no segundo tempo: 1038
```

```
jogadores_mais_punidos = cartoes_2023['atleta'].value_counts().sort_values(ascending=False)
jogadores_mais_punidos.head(10)
                         count
                atleta
   Walter Kannemann
    Gabriel Mercado
                            13
   Robson Fernandes
                            13
   Rodrigo Fernández
                            13
Reinaldo Manoel da Silva
                            12
    Gustavo Gómez
                            12
      Diego Costa
                            12
    Mauricio Lemos
                            12
   Juninho Capixaba
                            12
         André
                            12
dtype: int64
```