

# DATA SCIENCE & STATISTICAL COMPUTING

**LAB - Analisar Dados do Brasileirão**

Prof. MEng. Daniel de Souza Carvalho

[profdaniel.carvalho@fiap.com.br](mailto:profdaniel.carvalho@fiap.com.br)

@danielscarvalho

# LAB - Analisar Dados do Brasileirão

- **Parte 4**
- Carregar dados do “Brasileirão série A”
- Estabelecer questões de pesquisa, a classe toda
- Trabalhar em *paring*
- Montar as queries (consultas) no DataFrame Pandas para obter as respostas
- Utilizar referências já fornecidas (PDF, documentação, online), livros, etc...)
- Vamos trabalhar juntos no ambiente Jupyter



# LAB - Analisar Dados do Brasileirão

```
[1]: import pandas as pd
import numpy as np
```

```
[2]: !ls *.csv
```

```
brasileirao_serie_a.csv
```

```
[3]: bra = pd.read_csv("brasileirao_serie_a.csv")
```

```
[5]: bra.sample(2).T
```

```
[5]:
```

	3120	5973
ano_campeonato	2012	2015
data	2012-07-18	2015-08-09
rodada	10	17
estadio	Estádio Jornalista Mário Filho	Estádio Serra Dourada
arbitro	Sandro Ricci	Raphael Claus
publico	14897.0	7762.0
publico_max	NaN	50049.0
time_mandante	Flamengo	Goiás EC
time_visitante	Corinthians	Atlético-MG
tecnico_mandante	Joel Santana	Julinho Camargo

# LAB - Analisar Dados do Brasileirão

ano_campeonato	3120	5973	
data	2012	2015	
rodada	2012-07-18	2015-08-09	
estadio	10	17	
arbitro	Estádio Jornalista Mário Filho	Estádio Serra Dourada	
publico	Sandro Ricci	Raphael Claus	
publico_max	14897.0	7762.0	
time_mandante	NaN	50049.0	
time_visitante	Flamengo	Goiás EC	
tecnico_mandante	Corinthians	Atlético-MG	
tecnico_visitante	Joel Santana	Julinho Camargo	
colocacao_mandante	Tite	Levir Culpi	
colocacao_visitante	9.0	17.0	
valor_equipe_titular_mandante	15.0	1.0	
valor_equipe_titular_visitante	20100000.0	10500000.0	
idade_media_titular_mandante	21000000.0	28500000.0	
idade_media_titular_visitante	26.7	24.9	
gols_mandante	28.8	27.9	
gols_visitante	0.0	0.0	
gols_1_tempo_mandante	3.0	0.0	
gols_1_tempo_visitante	0.0	0.0	
escanteios_mandante	2.0	0.0	
escanteios_visitante	NaN	NaN	
faltas_mandante	NaN	NaN	
faltas_visitante	NaN	NaN	
chutes_bola_parada_mandante	NaN	NaN	
chutes_bola_parada_visitante	NaN	NaN	
defesas_mandante	NaN	NaN	
defesas_visitante	NaN	NaN	
impedimentos_mandante	NaN	NaN	
impedimentos_visitante	NaN	NaN	
chutes_mandante	NaN	NaN	
chutes_visitante	NaN	NaN	
chutes_fora_mandante	NaN	NaN	
chutes_fora_visitante	NaN	NaN	

# LAB - Analisar Dados do Brasileirão

```
[6]: bra.describe()
```



[6]:	ano_campeonato	rodada	publico	publico_max	colocacao_mandante	colocacao_visitante	valor_equipe_titular_mandante	valor_equipe_titular_visitante
count	8079.000000	8079.000000	6368.000000	3817.000000	6369.000000	6369.000000	5.981000e+03	5.981000e+03
mean	2012.670999	19.337294	15548.533134	43767.216662	10.630711	10.371016	1.716411e+07	1.673819e+07
std	6.082025	10.916278	13677.813939	19832.337668	5.756188	5.773987	1.458502e+07	1.396350e+07
min	2003.000000	1.000000	0.000000	5000.000000	1.000000	1.000000	5.600000e+02	5.930000e+02
25%	2007.000000	10.000000	5684.000000	24584.000000	6.000000	5.000000	6.800000e+06	6.800000e+06
50%	2013.000000	19.000000	12286.500000	43600.000000	11.000000	10.000000	1.270000e+07	1.255000e+07
75%	2018.000000	29.000000	21916.250000	57483.000000	16.000000	15.000000	2.470000e+07	2.375000e+07
max	2023.000000	38.000000	82044.000000	87101.000000	20.000000	20.000000	1.005000e+08	9.600000e+07

8 rows × 28 columns

# LAB - Analisar Dados do Brasileirão

Questões sobre os dados do “Brasileirão série A”:

Usar ambiente Jupyter notebook!!

## Tabela Brasileirão Série A

POSICÃO		PONTOS	POSICÃO		PONTOS
1º	 Botafogo	33	11º	 Fortaleza	20
2º	 Grêmio	26	12º	 Cruzeiro	18
3º	 Flamengo	25	13º	 Cuiabá	15
4º	 Palmeiras	23	14º	 Santos	13
5º	 Bragantino	23	15º	 Bahia	12
6º	 Fluminense	21	16º	 Corinthians	12
7º	 São Paulo	21	17º	 Goiás	11
8º	 Internacional	21	18º	 Vasco	9
9º	 Athletico-PR	20	19º	 América-MG	9
10º	 Atlético-MG	20	20º	 Coritiba	7

LIBERTADORES

PRÉ-LIBERTADORES

ZONA DE REBAIXAMENTO

# LAB - Analisar Dados do Brasileiro

