



# Campeonato Brasileiro

João Victor do Couto de Azevêdo  
Ciência da Computação - UENF

# Dados

A base de informações utilizadas neste estudo consistem em dados de partidas de futebol do campeonato brasileiro desde a edição de 2003, quando se iniciou o formato de pontos corridos na competição, e também de dados sobre o histórico de campeões e outras colocações do topo da tabela do campeonato brasileiro.

## Colunas

### **campeonato\_brasileiro**

- id: identificador único da partida
- rodada: identificador da rodada correspondente à partida que foi realizada
- data: data em que foi realizada a partida
- horario: horário em que foi realizada a partida
- dia: dia da semana em que foi realizada a partida
- mandante: nome da equipe mandante
- visitante: nome da equipe visitante
- vencedor: nome da equipe que venceu a partida
- arena: nome do estádio em que foi realizada a partida
- mandante\_placar: quantos gols a equipe mandante marcou
- visitante\_placar: quantos gols a equipe visitante marcou
- estado\_mandante: estado ao qual pertence a equipe mandante
- estado\_visitante: estado ao qual pertence a equipe visitante
- estado\_vencedor: estado ao qual pertence a equipe que venceu a partida
- ano\_brasileirao: ano da edição do campeonato que foi realizado

### **campeoes\_brasileiros**

- ano: ano da edição do campeonato que foi realizado
- vencedor: primeiro colocado na tabela de classificação após a última rodada
- vice: segundo colocado na tabela de classificação após a última rodada
- terceiro\_colocado: terceiro colocado na tabela de classificação após a última rodada
- quarto\_colocado: quarto colocado na tabela de classificação após a última rodada
- artilheiro: o(s) jogador(es) que marcou(marcaram) mais gols na edição do campeonato
- gols: quantos gols o(s) artilheiro(s) marcou

# Caso de Estudo

No futebol, em especial no futebol brasileiro, é dito que quando um time joga em casa, obtém vantagem em relação ao time visitante, por estar mais acostumado com a torcida e sofre menos hostilidade.

Em 2020, com a pandemia do coronavírus, foi realizado o campeonato sem público presente, a fim de evitar aglomerações.

O estudo será para analisar o aproveitamento dos times, e ver como este aproveitamento pode influenciar na colocação final, ou até em título. Em especial analisar o aproveitamento durante a edição de 2020.

Como o campeonato brasileiro só foi ter o formato atual em 2006, com 20 times, vamos utilizar uma amostra entre 2006 e 2019.

```
campeoes_brasileiros_filtrados = campeoes_brasileiros |>
+   filter(Ano >= 2006 & Ano < 2022)

> View(campeoes_brasileiros_filtrados)
```

	Ano	Vencedor	Vice	Terceiro colocado	Quarto colocado	Artilheiro(s)	Gols
1	2006	São	Internacional	Grêmio	Santos	Souza (Goiás)	17
2	2007	São	Santos	Flamengo	Fluminense	Josiel (Paraná)	20
3	2008	São	Grêmio	Cruzeiro	Palmeiras	Keirison (Coritiba) Kléber Pereira (Santos) Washington (Flu...	21
4	2009	Flamengo	Internacional	São Paulo	Cruzeiro	Adriano (Flamengo) Diego Tardelli (Atlético Mineiro)	19
5	2010	Fluminense	Cruzeiro	Corinthians	Grêmio	Jonas (Grêmio)	23
6	2011	Corinthians	Vasco da Gama	Fluminense	Flamengo	Borges (Santos)	23
7	2012	Fluminense	Atlético Mineiro	Grêmio	São Paulo	Fred (Fluminense)	20
8	2013	Cruzeiro	Grêmio	Atlético Paranaense	Botafogo	Éderson (Atlético Paranaense)	21
9	2014	Cruzeiro	São Paulo	Internacional	Corinthians	Fred (Fluminense)	18
10	2015	Corinthians	Atlético Mineiro	Grêmio	São Paulo	Ricardo Oliveira (Santos)	20
11	2016	Palmeiras	Santos	Flamengo	Atlético Mineiro	William Pottker (Ponte Preta) Diego Souza (Sport) Fred (At...	14
12	2017	Corinthians	Palmeiras	Santos	Grêmio	Henrique Dourado (Fluminense) Jô (Corinthians)	18
13	2018	Palmeiras	Flamengo	Internacional	Grêmio	Gabriel (Santos)	18
14	2019	Flamengo	Santos	Palmeiras	Grêmio	Gabriel (Flamengo)	25

Como podemos já podemos notar, o maior artilheiro deste tipo de edição com 20 clubes, é o Gabigol com 25 gols.

Para filtrarmos apenas os jogos apartir de 2016 e criarmos uma nova coluna com o ano daquele jogo:

```
jogos_brasileirao_filtrados = jogos_brasileirao |>
+   filter(substr(data, 7, 10) >= 2006 & substr(data, 7, 10) < 2020) |>
+   mutate(ano_brasileirao = substr(data, 7, 10))
```

Para podermos visualizar melhor, é necessário criarmos duas novas tabelas, com os gols do visitantes e dos mandantes

```
df_mandantes_gols = jogos_brasileirao_filtrados |>
  group_by(ano_brasileirao) |>
  summarise(gols = sum(mandante_Placar)) |>
  mutate(tipo_gols = "gols dos mandantes")
```

ano_brasileirao	gols	tipo_gols
2006	91	gols dos mandantes
2007	95	gols dos mandantes
2008	113	gols dos mandantes
2009	134	gols dos mandantes
2010	110	gols dos mandantes
2011	117	gols dos mandantes
2012	116	gols dos mandantes
2013	129	gols dos mandantes
2014	117	gols dos mandantes
2015	99	gols dos mandantes
2016	109	gols dos mandantes
2017	131	gols dos mandantes
2018	109	gols dos mandantes
2019	119	gols dos mandantes

```
df_visitantes_gols = jogos_brasileirao_filtrados %>%
  group_by(ano_brasileirao) %>%
  summarise(gols = sum(visitante_Placar)) %>%
  mutate(tipo_gols = "gols dos visitantes")
```

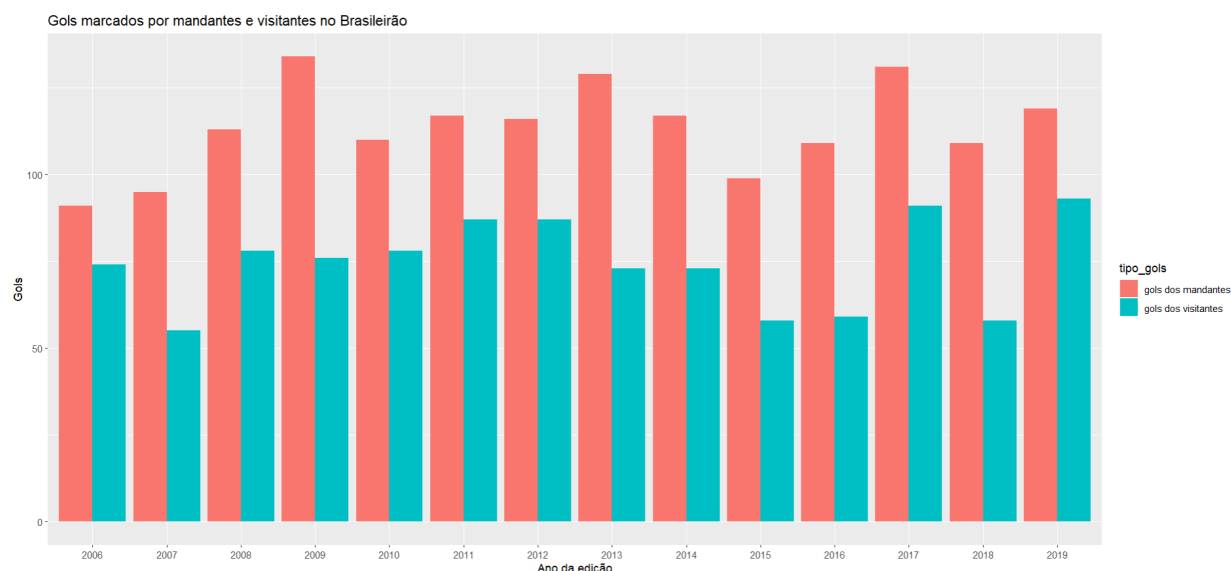
ano_brasileirao	gols	tipo_gols
2006	74	gols dos visitantes
2007	55	gols dos visitantes
2008	78	gols dos visitantes
2009	76	gols dos visitantes
2010	78	gols dos visitantes
2011	87	gols dos visitantes
2012	87	gols dos visitantes
2013	73	gols dos visitantes
2014	73	gols dos visitantes
2015	58	gols dos visitantes
2016	59	gols dos visitantes
2017	91	gols dos visitantes
2018	58	gols dos visitantes
2019	93	gols dos visitantes

Para criarmos uma tabela com todos os gols:

```
df_gols = rbind(df_mandantes_gols, df_visitantes_gols)
```

Para visualizar melhor os dados, iremos analisar um gráfico entre os gols marcados por visitantes e mandantes.

```
df_gols %>%  
+   ggplot(aes(fill=tipo_gols, y=gols, x=as.character(ano_brasileirao))) +  
+   geom_bar(position="dodge", stat="identity") +  
+   labs(  
+     x = "Ano da edição",  
+     y = "Gols",  
+     title = "Gols marcados por mandantes e visitantes no Brasileirão"  
+   )
```



Em todos os anos, desde 2006, podemos observar que o time mandante marcou mais gols. Porém, para fazer uma melhor análise, precisamos ver o rendimento dos times, vitórias e derrotas.

Para criarmos uma tabela com o número de vitória dos mandantes:

```
df_vitorias_mandantes = jogos_brasileirao_filtrados %>%  
  group_by(ano_brasileirao) %>%  
  summarise(  
    jogos = sum(vencedor == mandante)  
  ) %>%  
  mutate(tipo_jogos = "1 - Vitória dos mandantes")
```

	ano_brasileirao	jogos	tipo_jogos
1	2006	28	1 - Vitória dos mandantes
2	2007	32	1 - Vitória dos mandantes
3	2008	30	1 - Vitória dos mandantes
4	2009	43	1 - Vitória dos mandantes
5	2010	34	1 - Vitória dos mandantes
6	2011	39	1 - Vitória dos mandantes
7	2012	37	1 - Vitória dos mandantes
8	2013	47	1 - Vitória dos mandantes
9	2014	40	1 - Vitória dos mandantes
10	2015	33	1 - Vitória dos mandantes
11	2016	38	1 - Vitória dos mandantes
12	2017	40	1 - Vitória dos mandantes
13	2018	43	1 - Vitória dos mandantes
14	2019	40	1 - Vitória dos mandantes

O mesmo para Vitória dos Visitantes e empates:

```
df_vitorias_visitantes = jogos_brasileirao_filtrados |>
  group_by(ano_brasileirao) |>
  summarise(jogos = sum(vencedor == visitante)) |>
  mutate(tipo_jogos = "2 - Vitória dos visitantes")
```

```
df_empates = jogos_brasileirao_filtrados |>
  group_by(ano_brasileirao) |>
  summarise(jogos = sum(vencedor == '-')) |>
  mutate(tipo_jogos = "3 - Empates")
```

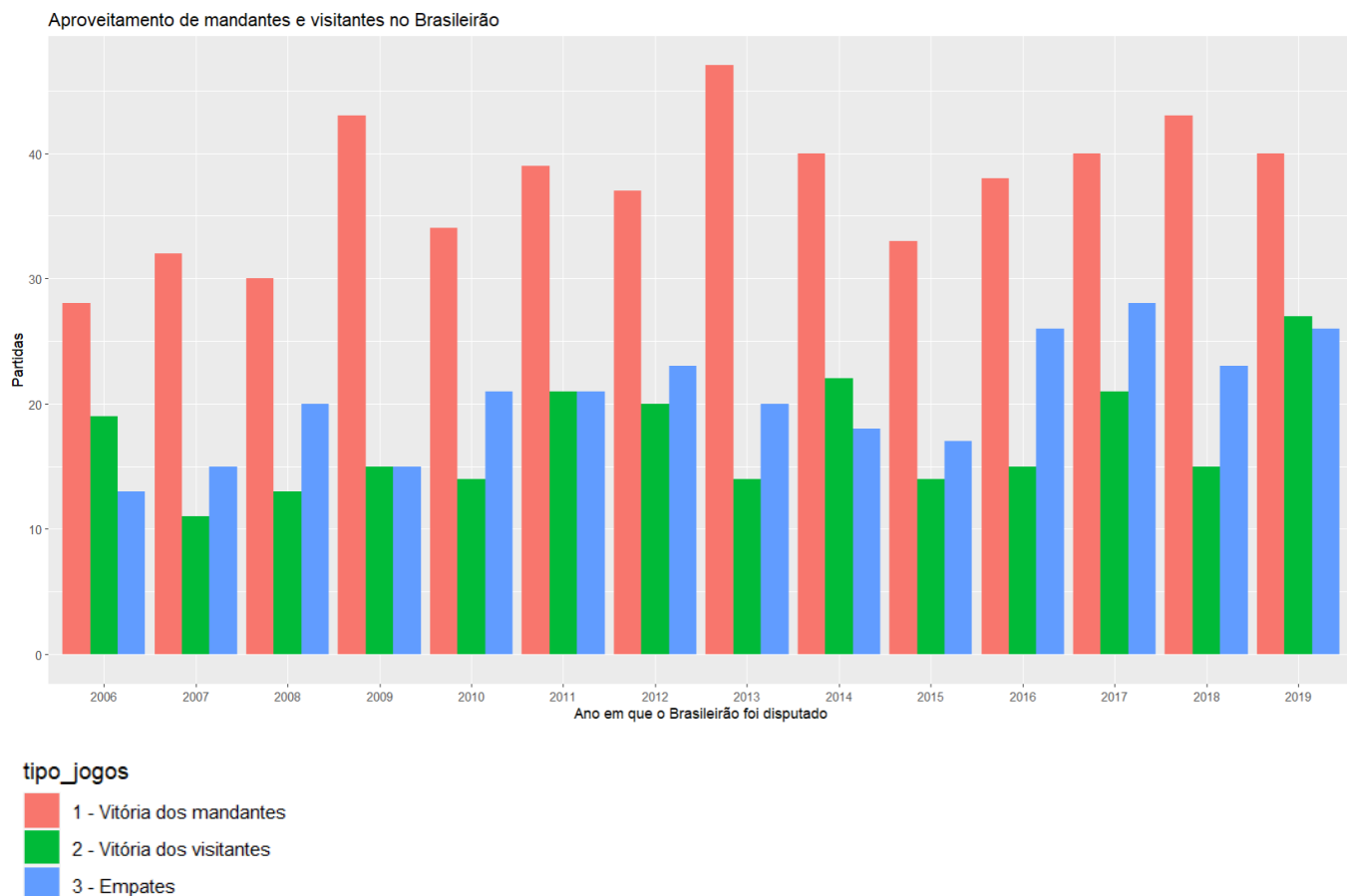
Para termos uma com todos os jogos:

```
df_jogos = rbind(df_vitorias_mandantes, df_vitorias_visitantes, df_empates)
```

11	2016	38	1 - Vitória dos mandantes
12	2017	40	1 - Vitória dos mandantes
13	2018	43	1 - Vitória dos mandantes
14	2019	40	1 - Vitória dos mandantes
15	2006	19	2 - Vitória dos visitantes
16	2007	11	2 - Vitória dos visitantes
17	2008	13	2 - Vitória dos visitantes
18	2009	15	2 - Vitória dos visitantes
19	2010	14	2 - Vitória dos visitantes
20	2011	21	2 - Vitória dos visitantes
21	2012	20	2 - Vitória dos visitantes
22	2013	14	2 - Vitória dos visitantes
23	2014	22	2 - Vitória dos visitantes
24	2015	14	2 - Vitória dos visitantes
25	2016	15	2 - Vitória dos visitantes
26	2017	21	2 - Vitória dos visitantes
27	2018	15	2 - Vitória dos visitantes
28	2019	27	2 - Vitória dos visitantes
29	2006	13	3 - Empates
30	2007	15	3 - Empates
31	2008	20	3 - Empates
32	2009	15	3 - Empates
33	2010	21	3 - Empates
34	2011	21	3 - Empates
35	2012	23	3 - Empates
36	2013	20	3 - Empates
37	2014	18	3 - Empates
38	2015	17	3 - Empates

Para obtermos uma melhor visualização do aproveitamento desses times, iremos utilizar um gráfico:

```
df_jogos %>%
  ggplot(aes(fill=tipo_jogos, y=jogos, x=as.character(ano_brasileirao)))
  geom_bar(position="dodge", stat="identity")
  labs(
    x = "Ano em que o Brasileirão foi disputado",
    y = "Partidas",
    title = "Aproveitamento de mandantes e visitantes no Brasileirão")
```



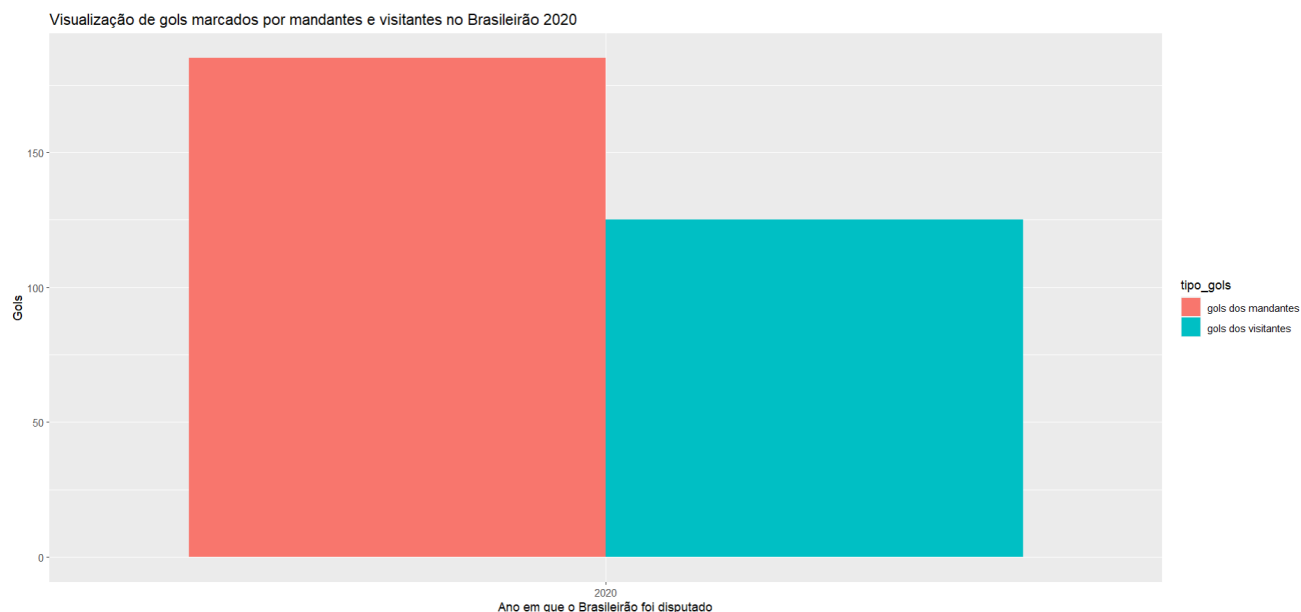
De acordo com a análise do gráfico, é possível vermos que os mandantes costumam ganhar mais jogos do que os visitantes. O empate também se mostra em uma concorrência maior com as vitórias de visitantes.

Com a adversidade do coronavírus, o brasileirão de 2020 sofreu algumas alterações e algumas obrigações para jogar, como o de jogar com os estádios vazios, afetando o desempenho deles.

Esta análise foi feita com o objetivo de ver claramente se o fator mandante no campeonato brasileiro tem sua vantagem, e como isso foi afetado no ano em que não tivemos presença nos estádios.

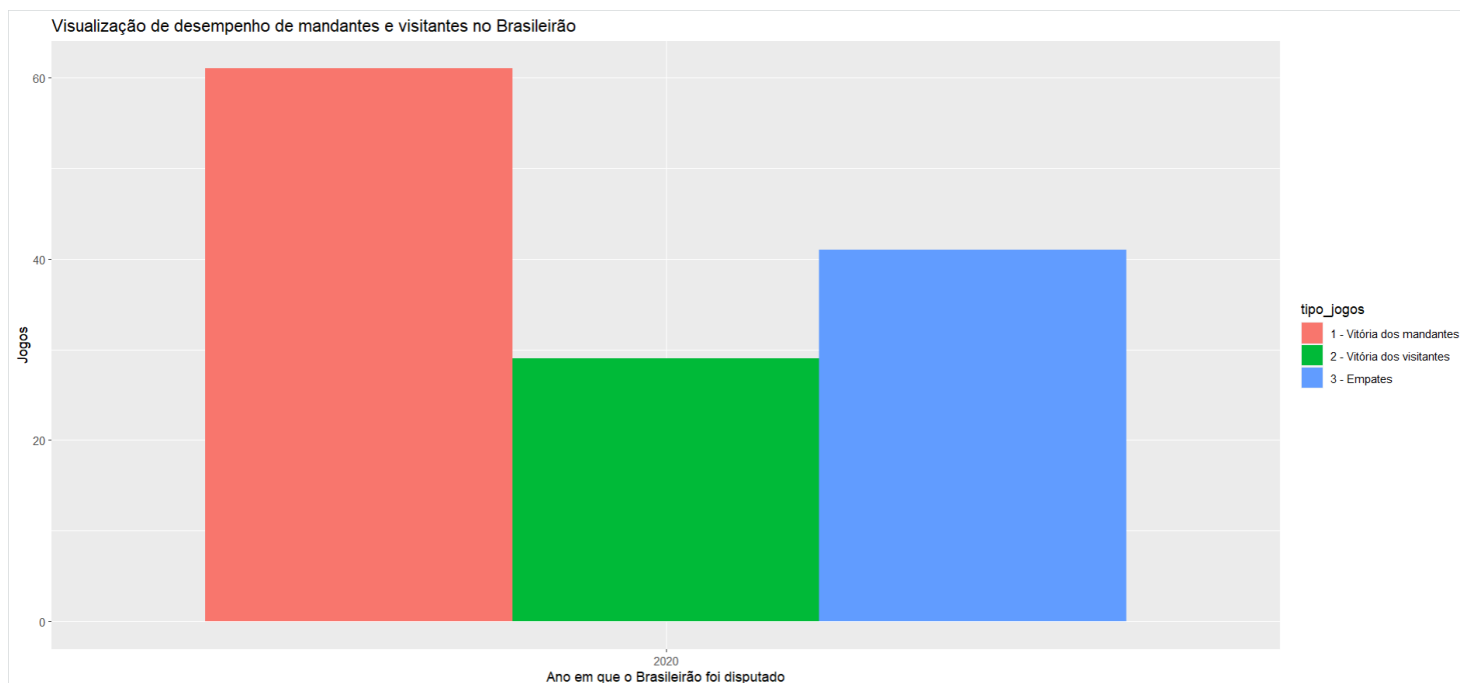
Outra análise que poderia surgir é sobre os confrontos no mata-mata, a forma que a competição era jogada antes dos pontos corridos.





(Os números não ficaram corretos, porém a proporção do gráfico está correta)

Continua uma vantagem dos times mandantes em relação aos que jogam fora de casa, já que além destes aspectos, como o Brasil é um país muito grande com muita diversidade, alguns confrontos precisam ser disputado muito longe de sua cidade, realizando uma viagem cansativa e chegando a um clima completamente diferente do que está acostumado, por exemplo um time da Bahia ir a Porto Alegre para jogar.



Devido a esses fatores, os mandantes continuaram tendo uma certa vantagem ao jogar em casa, porém sem a pressão da torcida adversária conseguiram levemente diminuir a vantagem.