

# Analise Exploratória do Campeonato Brasileiro de Futebol (2003-2023)

Nikolas N. Figueiredo

11 de junho de 2024

## Contents

<b>Introdução</b>	<b>1</b>
<b>Objetivos do Relatório</b> . . . . .	1
<b>Pacotes Utilizados</b> . . . . .	2
<b>Primeiro Contato com os Dados</b>	<b>2</b>
<b>Análise dos Dados</b>	<b>3</b>
<b>Observações Gerais</b> . . . . .	3
<b>Observações em Relação ao Ano</b> . . . . .	5
<b>Conclusões</b>	<b>5</b>
<b>Referências</b>	<b>5</b>

---

## Introdução

Este relatório surge como um estudo do Campeonato Brasileiro Série A, que é um dos objetos de estudo da minha dissertação de mestrado em Estatística Aplicada e Biometria. As análises foram feitas em linguagem R, através da IDE RStudio e o relatório foi gerado em R Markdown. Os dados utilizados foram disponibilizados por DUQUE (2023), por meio da plataforma Kaggle.

## Objetivos do Relatório

O objetivo deste relatório é apresentar uma análise exploratória dos dados do Campeonato Brasileiro Série A, com foco em identificar padrões, tendências e insights que possam ser úteis para análises futuras e para a tomada de decisões estratégicas.

## Objetivos Específicos

Ainda deseja-se:

- Observar quais times participaram dessa competição.
- Contabilizar a participação dos times e sua relevância para a competição.
- Entender a distribuição dos gols na competição e em cada ano do campeonato.
- Visualizar a distribuição dos jogos em relação aos estádios e aos estados do Brasil.

## Pacotes Utilizados

```
#Pacotes
library(tidyverse, plotly)
```

## Primeiro Contato com os Dados

Leitura dos dados.

```
br <- read_csv("br.csv")
## Rows: 8405 Columns: 17
## -- Column specification -----
## Delimiter: ","
## chr (11): data, mandante, visitante, formacao_mandante, formacao_visitante,...
## dbl (5): ID, rodata, ano, mandante_Placar, visitante_Placar
## time (1): hora
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Observando a estrutura dos dados, é possível visualizar uma tabela com **8,405** linhas e **17** colunas. Cada uma dessas linhas estão representando uma partida, assim contém: informações da data, hora, local da partida; informações dos times que se enfrentaram, a formação técnica e o técnico responsável; e informações sobre o resultado do confronto como placar e time vencedor.

```
## Rows: 8,405
## Columns: 17
## $ ID <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ~
## $ rodata <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, ~
## $ data <chr> "29/03/2003", "29/03/2003", "30/03/2003", "30/03/20~
## $ ano <dbl> 2003, 2003, 2003, 2003, 2003, 2003, 2003, 2003, 200~
## $ hora <time> 16:00:00, 16:00:00, 16:00:00, 16:00:00, 16:00:00, ~
## $ mandante <chr> "Guarani", "Athletico-PR", "Flamengo", "Goias", "In~
## $ visitante <chr> "Vasco", "Gremio", "Coritiba", "Paysandu", "Ponte P~
## $ formacao_mandante <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ formacao_visitante <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ tecnico_mandante <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ tecnico_visitante <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ vencedor <chr> "Guarani", "Athletico-PR", "-", "-", "-", "Criciuma~
## $ arena <chr> "Brinco de Ouro", "Arena da Baixada", "Maracanã"~
## $ mandante_Placar <dbl> 4, 2, 1, 2, 1, 2, 2, 0, 2, 1, 2, 0, 1, 0, 0, 3, 1, ~
## $ visitante_Placar <dbl> 2, 0, 1, 2, 1, 0, 2, 0, 2, 1, 2, 3, 1, 0, 1, 1, 2, ~
## $ mandante_Estado <chr> "SP", "PR", "RJ", "GO", "RS", "SC", "RS", "CE", "MG~
## $ visitante_Estado <chr> "RJ", "RS", "PR", "PA", "SP", "RJ", "SP", "BA", "SP~
```

**Observação:** Ao contrário do que o comando acima sugere, as colunas de formação tática e do técnico responsável pelos times mandante e visitante, não são colunas inteiramente compostas por “NA”. As informações sobre o técnico começaram a ser coletadas a partir de 2014 e em 2015 foram adicionadas informações sobre a formação dos times. Como essas variáveis não têm impacto significativo nos objetivos da análise a abordagem adotada foi: Manter os dados existentes, não adotar nenhuma técnica de completamento dos dados faltantes e apenas observar essas métricas a partir dos anos indicados.

```
sum(table(br$formacao_mandante))
## [1] 3431
```

```
sum(table(br$formacao_visitante))
## [1] 3431
```

```
sum(table(br$tecnico_mandante))
## [1] 3795
```

```
sum(table(br$tecnico_visitante))
## [1] 3795
```

## Analise dos Dados

Esta análise está dividida em duas perspectivas, primeiro de forma geral considerando todos os anos de campeonato e, segundo, as métricas serão observadas em relação a cada ano de maneira separada.

### Observações Gerais

Estão representados os campeonatos de 2003 a 2023, totalizando 21 anos de campeonato.

```
## # A tibble: 21 x 4
##   Ano `Número de jogos` `Número de rodadas` `Número de times`
##   <dbl>          <int>          <dbl>          <dbl>
## 1 2003             552             46             24
## 2 2004             552             46             24
## 3 2005             462             42             22
## 4 2006             380             38             20
## 5 2007             380             38             20
## 6 2008             380             38             20
## 7 2009             380             38             20
## 8 2010             380             38             20
## 9 2011             380             38             20
## 10 2012            380             38             20
## 11 2013            380             38             20
## 12 2014            380             38             20
## 13 2015            380             38             20
## 14 2016            379             38             20
## 15 2017            380             38             20
## 16 2018            380             38             20
## 17 2019            380             38             20
## 18 2020            380             38             20
## 19 2021            380             38             20
## 20 2022            380             38             20
## 21 2023            380             38             20

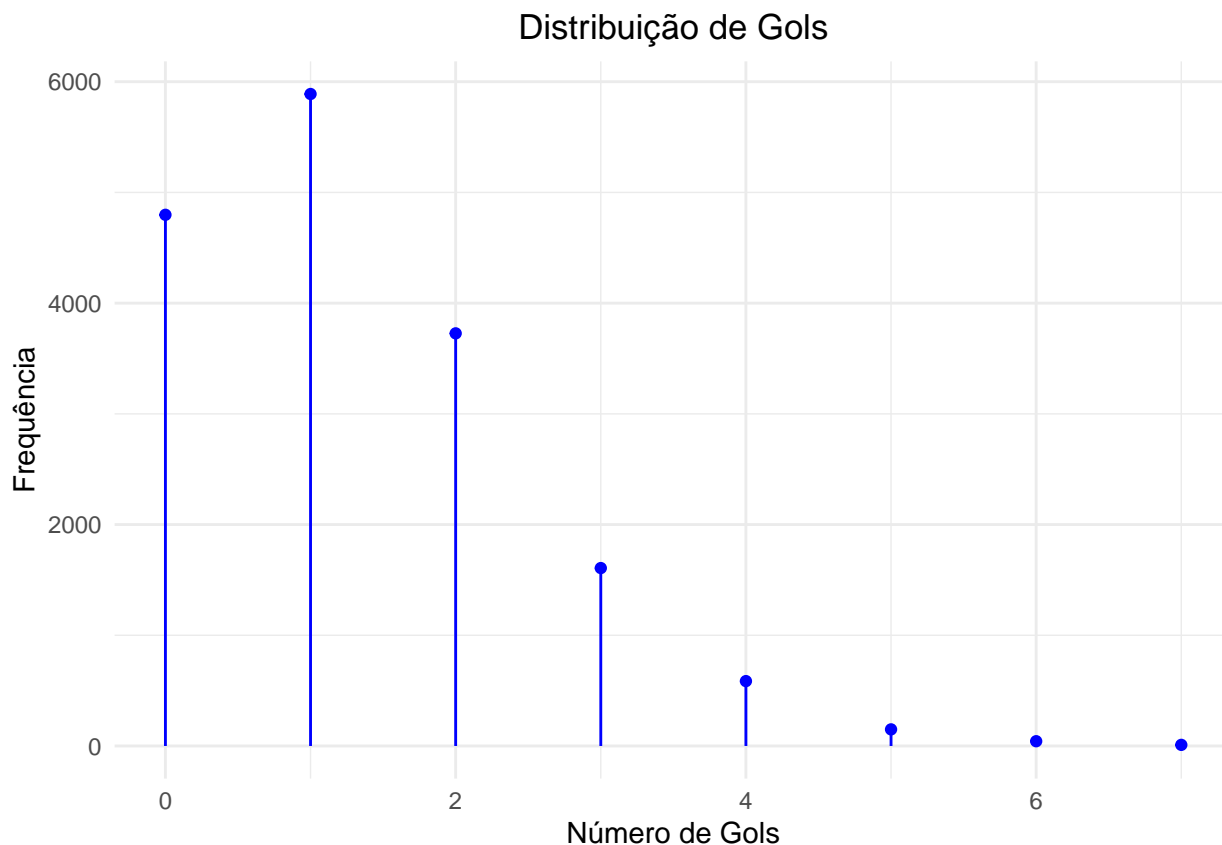
##           Time N° jogos N° participações Vitórias Empates Derrotas
## 1   America-MG    228           6         58      29      109
## 2   America-RN     38           1          4       3       29
## 3 Athletic-PR    780          20        309     102     282
## 4 Atletico-GO   266           7         72      41     115
## 5 Atletico-MG   779          20        327      83     253
## 6      Avai     266           7         66      45     127
## 7      Bahia    426          11        127      64     178
## 8    Barueri     38           1         12       7      13
## 9 Botafogo-RJ   696          18        247      93     259
```

## 10	Bragantino	152	4	55	24	47
## 11	Brasiliense	42	1	10	7	20
## 12	Ceara	266	7	72	48	102
## 13	Chapecoense	265	7	70	37	118
## 14	Corinthians	780	20	326	109	225
## 15	Coritiba	590	15	189	73	250
## 16	Criciuma	168	4	50	18	80
## 17	Cruzeiro	704	18	304	87	235
## 18	CSA	38	1	8	5	22
## 19	Cuiaba	114	3	34	20	43
## 20	Figueirense	438	11	142	69	172
## 21	Flamengo	818	21	354	104	243
## 22	Fluminense	818	21	322	111	283
## 23	Fortaleza	316	8	108	41	129
## 24	Goiias	590	15	209	78	239
## 25	Gremio	738	19	316	85	237
## 26	Gremio Prudente	38	1	7	5	21
## 27	Guarani	130	3	36	22	55
## 28	Internacional	780	20	339	93	239
## 29	Ipatinga	38	1	9	3	21
## 30	Joinville	38	1	7	7	21
## 31	Juventude	286	7	85	51	125
## 32	Nautico	190	5	54	19	98
## 33	Palmeiras	734	19	329	79	215
## 34	Parana	248	6	83	30	110
## 35	Paysandu	134	3	41	16	62
## 36	Ponte Preta	362	9	114	40	157
## 37	Portuguesa	114	3	31	23	45
## 38	Santa Cruz	76	2	15	6	47
## 39	Santo Andre	38	1	11	5	19
## 40	Santos	818	21	340	106	261
## 41	Sao Caetano	172	4	65	20	66
## 42	Sao Paulo	818	21	365	111	224
## 43	Sport	418	11	131	58	180
## 44	Vasco	628	16	208	91	234
## 45	Vitoria	396	10	123	53	174

Nesses 21 anos de competição foram feitos 21.586 gols que foram distribuidos da seguinte de 0 a 7.

##	gols	Freq
## 1	0	4798
## 2	1	5889
## 3	2	3727
## 4	3	1607
## 5	4	586
## 6	5	150
## 7	6	43
## 8	7	10

## Número total de gols = 21586  
## A média de gols = 1.284117  
## O desvio padrão = 1.162225



### Observações em Relação ao Ano

---

### Conclusões

---

### Referências

- R CORE TEAM (2024)
- RSTUDIO TEAM (2023)
- XIE; ALLAIRE; GROLEMUND (2018)

DUQUE, A. **Campeonato brasileiro de futebol**. [S. l.]: Kaggle, 2023. Available at: <https://www.kaggle.com/datasets/adaoduke/campeonato-brasileiro-de-futebol>.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2024. Available at: <https://www.R-project.org/>.

RSTUDIO TEAM. **RStudio: Integrated development environment for r**. Boston, MA: RStudio, PBC, 2023. Available at: <http://www.rstudio.com/>.

XIE, Y.; ALLAIRE, J.; GROLEMUND, G. **R markdown: The definitive guide**. [S. l.: s. n.], 2018. Available at: <https://bookdown.org/yihui/rmarkdown>.