

Dataset público sobre times e partidas do Campeonato Brasileiro de Futebol

Equipe AGoldenDatabase - Trabalho Final MC536 2°
Semestre 2021

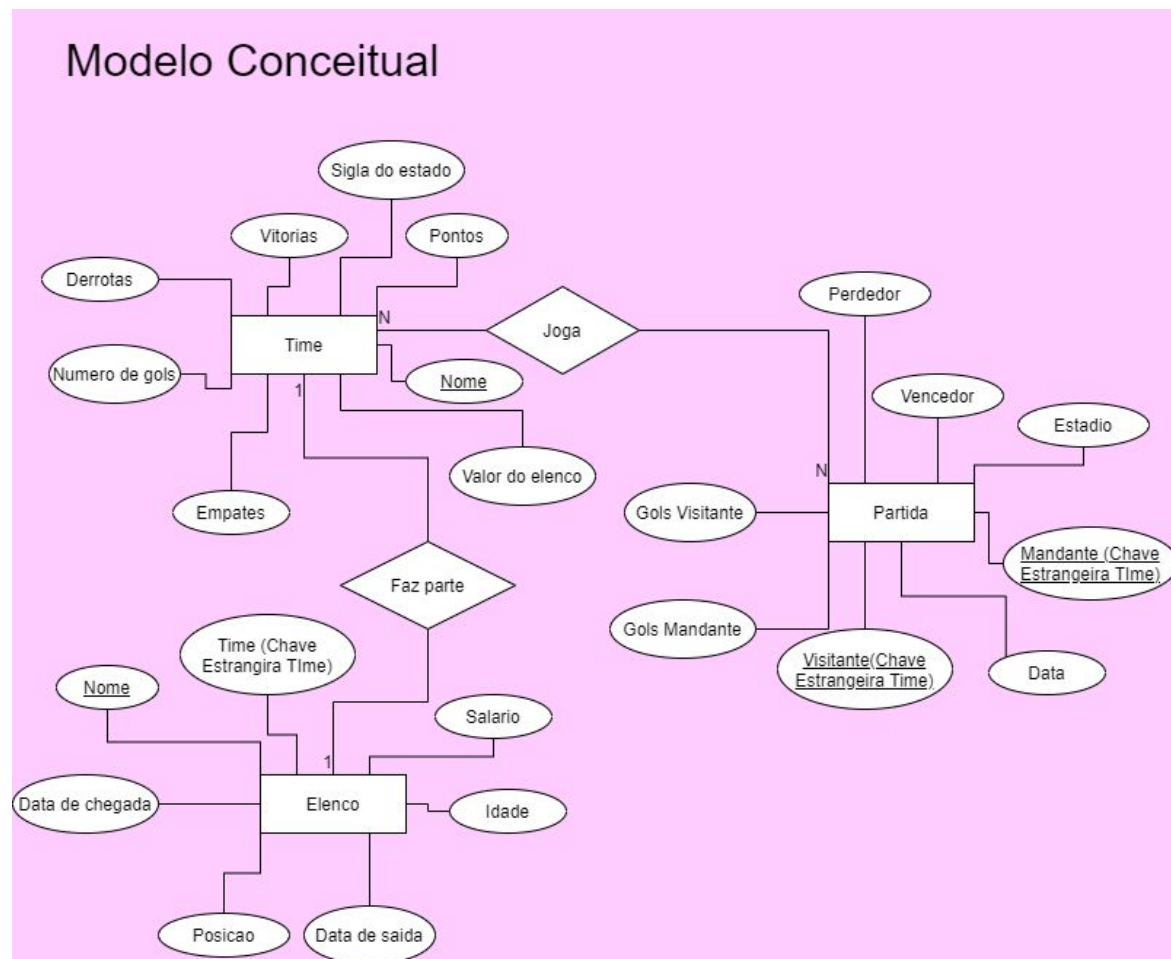
Motivação e Contexto Gerador

Ultimamente tem se tornado muito comum a análise de eventos esportivos orientada a dados, especialmente para realizar previsões e apostas de curto e longo prazo sobre os resultados de um campeonato. No entanto, muitos sites de estatística ainda carecem de informações que podem ser relevantes para o entendimento dos resultados do campeonato. Nesse sentido, buscamos compilar e formatar dados diversos sobre os times e partidas no Brasileirão em um banco de dados que permita traçar correlações entre as condições dos times e seus resultados.

Perguntas a serem respondidas

- Qual a chance de um time ganhar, dado seu histórico e dados sobre seu elenco?
- Qual a correlação entre valor de mercado ou salário do elenco e seu desempenho em um campeonato?
- Qual o efeito da idade média dos jogadores no resultado dos times?

Modelo Conceitual



Modelos Lógicos

- Optamos por utilizar dois modelos lógicos para acesso ao nosso dataset:
- O primeiro será organizado em um modelo tabular, em formato csv, separados em arquivos diferentes seguindo a lógica do modelo conceitual.
- O segundo, será disponibilizado em forma semi-estruturada (documentos), no formato xml, contendo todas as informações registradas a partir do nó principal.

Fonte de dado candidata (1)

[Sofascore:](#)

- Jogadores de cada time
- Gols Marcados
- Número de assistências
- Posição
- Idade

SofaScore

Search

🔍

📉 Dropping odds

🔔 Favorites

👤 Profile

1/438

105

2/174

9/475

138

2/40

55

1/56

29

4

📺

18

...

⚙️

Football

Brazil

Brasileirão Série A live score

Brasileirão Série A

🇧🇷 Brazil

2021 ▾

Receive notifications for all games in this tournament

FOLLOW 📌

203K followers

29 MAY

5 DEC

STANDINGS

OVERALL HOME AWAY

		P	W	D	L	Goals	Last 5	Pts
1	Atlético MG	21	14	4	3	32:13	D D W W D	46
2	Palmeiras	21	12	2	7	33:25	L W L W L	38
3	Fortaleza	22	10	6	6	30:23	D L L L W	36
4	Flamengo	19	11	2	6	36:19	D W W L D	35
5	RB Bragantino	21	8	9	4	32:24	W D L D L	33
6	Corinthians	22	8	9	5	22:19	W D D D W	33
7	Internacional	21	8	8	5	26:22	D D W W W	32
8	Fluminense	22	8	8	6	24:24	D W W D W	32
9	Athletico PR	21	9	3	9	29:26	L D L L W	30
10	Cuiabá	22	6	11	5	23:22	D W W D D	29
11	Ceará	21	6	10	5	20:21	D L L D W	28
12	Atlético GO	21	6	9	6	17:20	D D D L D	27
13	São Paulo	22	6	9	7	18:23	D L W D D	27
14	Juventude	22	6	8	8	21:25	D D L L W	26
15	América MG	22	5	9	8	19:24	W W D D D	24
16	Santos	22	5	9	8	20:28	L L D D L	24
17	Bahia	22	6	5	11	25:35	L W D D L	23
18	Grêmio	20	6	4	10	17:22	W L W W L	22
19	Sport Recife	22	3	8	11	8:19	D L L L L	17

TOP PLAYERS

Rating ▾

#			Rating
1	Lucas Crispim	Midfielder	7.37
2	Edenilson	Midfielder	7.36
3	Gustavo Scarpa	Midfielder	7.36
4	Artur	Midfielder	7.33
5	Juninho Valoura	Midfielder	7.32
6	Nikão	Midfielder	7.31
7	Daniel	Goalkeeper	7.24
8	Léo Ortiz	Defender	7.23
9	David Terans	Forward	7.23
10	Hulk	Forward	7.23

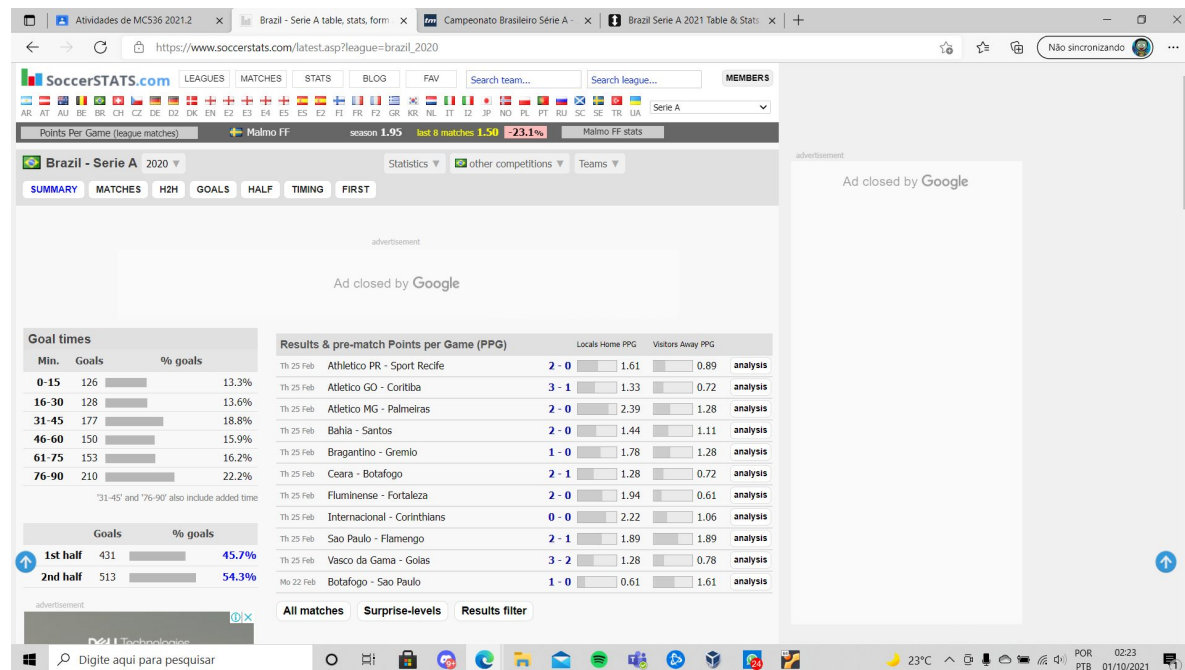
[Show more](#)

ADVERTISEMENT

Fonte de dado candidata (2)

[Brazil - Serie A table, stats, form and results | SoccerSTATS.com - historical 2020](https://www.soccerstats.com/latest.asp?league=brazil_2020)

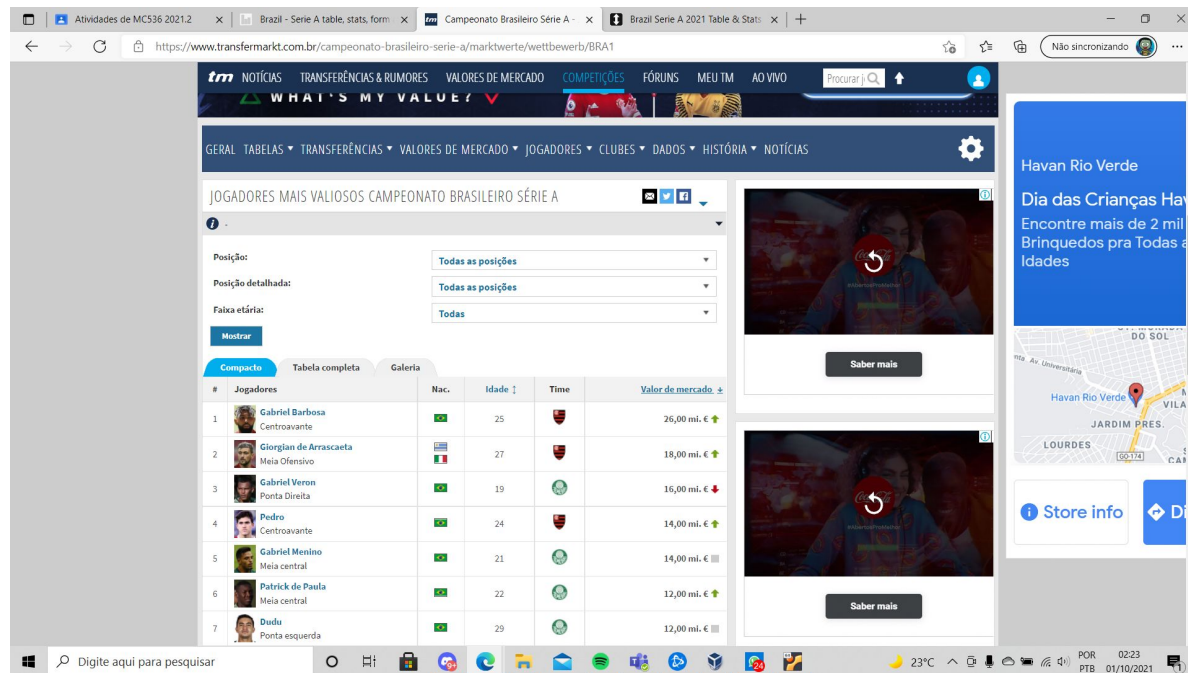
- Data das partidas
- Número de gols
- Placar
- Número de vitórias
- Número de derrotas
- UF de cada time



Fonte de dado Candidata (3)

Campeonato Brasileiro Série A - Jogadores mais valiosos 2021 | Transfermarkt

- Jogadores de cada time
- Salário de cada jogador
- Valor de cada jogador
- Valor investido de cada clube



#	Jogadores	Nac.	Idade	Time	Valor de mercado
1	Gabriel Barbosa Centroavante	BRA	25	Flamengo	26,00 mi. € ↑
2	Giorgian de Arrascaeta Meia Ofensivo	URU	27	Flamengo	18,00 mi. € ↑
3	Gabriel Veron Ponta Direita	BRA	19	Flamengo	16,00 mi. € ↓
4	Pedro Centroavante	BRA	24	Flamengo	14,00 mi. € ↑
5	Gabriel Menino Meia central	BRA	21	Flamengo	14,00 mi. €
6	Patrick de Paula Meia central	BRA	22	Flamengo	12,00 mi. € ↑
7	Dudu Ponta esquerda	BRA	29	Flamengo	12,00 mi. €

Procedimentos realizados sobre os dados

- Até o presente momento as fontes selecionadas não são estruturadas e portanto a coleta de dados será feita manualmente.
- Os dados ainda serão integrados de múltiplas fontes (citadas anteriormente)
- Outras estratégias poderão ser usadas futuramente, como a exploração de APIs relacionadas ao tema e uso de web scraping.