

Apresentação Final

Projeto de Banco de Dados - MC536



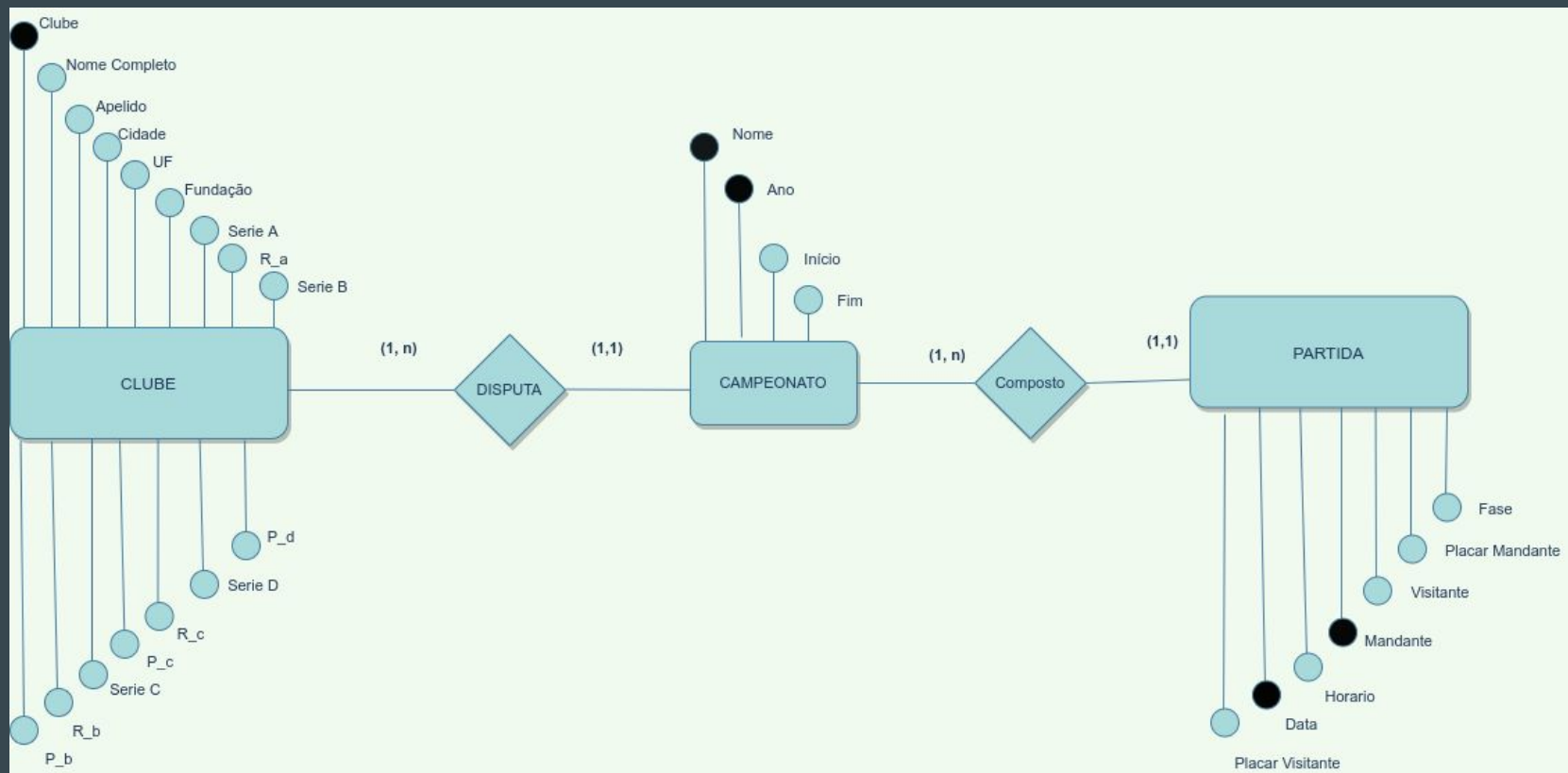
Grupo Dupla de Dois (DDD)
Gustavo Mantellatto Elias - 169366
Leandro Garcia Pereira - 178258

Campeonato Brasileiro Consolidado

Conforme exposto anteriormente, decidimos por criar um dataset contendo todas as partidas de Campeonato Brasileiro desde o seu início, em 1959. Isso foi motivado principalmente por conta da escassez de dados do tipo, visto que nem a própria Confederação Brasileira de Futebol o fez, para as primeiras edições.

Com isso, consolidamos esse banco de dados que nos permite uma enorme quantidade de análises e observações curiosas sobre os clubes brasileiros.

Modelo Conceitual

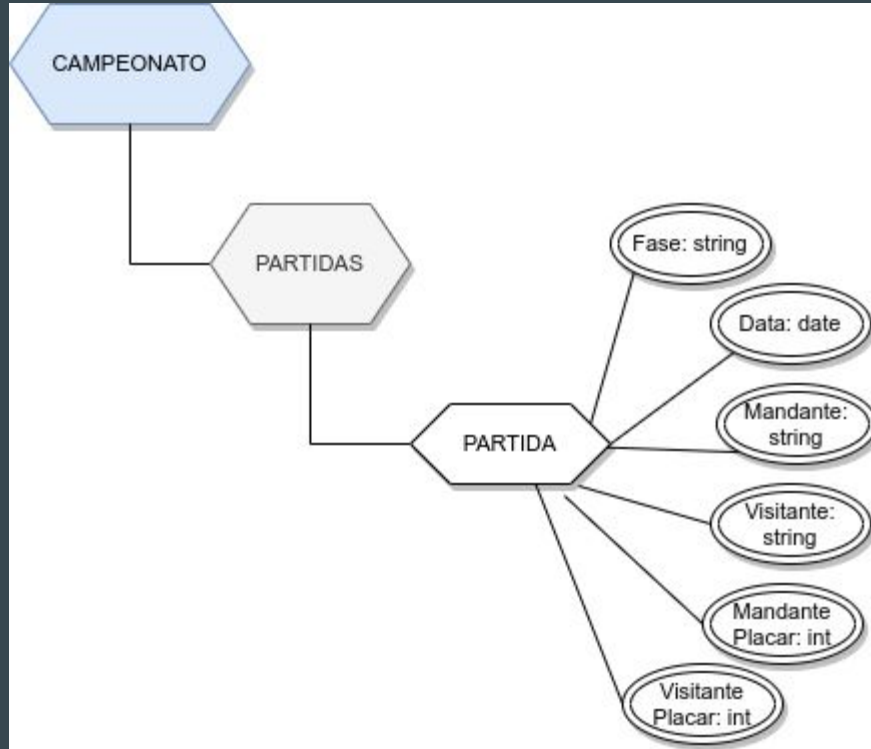


Modelo Relacional

```
CLUBE(_Clube_, Nome Completo, Cidade, UF, Fundação, Serie A, R_a, Serie B, P_b, R_b, Serie C, P_c, R_c, Serie D, P_d)
PARTIDA(_Data_, _Mandante_, Horario, Visitante, Placar_mandante, Placar_visitante)
    Mandante chave estrangeira -> Clube,
    Visitante chave estrangeira -> Clube
Campeonato(_Nome_, _Ano_, Inicio, Fim)
```



Modelo Hierárquico



Ferramentas

Toda a análise da parte que envolve nossas tabelas foi feita usando a ferramenta pandas, desde a extração com Web Scraping até as partes relacionadas a análise de dados e queries, conforme exemplo:

```
gols_mandante_clube = df.groupby('mandante')['placar_mandante'].sum().sort_values(ascending=False).reset_index()
gols_mandante_clube.rename(columns = {"mandante" : "clube", 'placar_mandante' : 'gols' }, inplace=True)

gols_visitante_clube = df.groupby('visitante')['placar_visitante'].sum().sort_values(ascending=False).reset_index()
gols_visitante_clube.rename(columns = {"visitante": "clube", 'placar_visitante': 'gols' }, inplace=True)
```

Análises

Conforme citado na última apresentação, um dos motivos interessantes de se ter um banco de dados com todas as partidas é poder fazer análises mais gerais de como os clubes do futebol brasileiro se comportam.

Por exemplo: é possível descobrir a proporção entre gols de mandantes e gols de visitantes em toda a história, como mostra a tabela a seguir:

	gols mandantes	gols visitantes	gols totais	gols mandantes perc	gols visitantes perc
0	31146	20325	51471	60.511744	39.488256

Análises

Ainda falando sobre os gols, temos como saber, por exemplo, quais times tem a maior quantidade de gols marcados e gols sofridos durante toda a história do campeonato.

Para o caso dos gols marcados, temos o seguinte resultado:

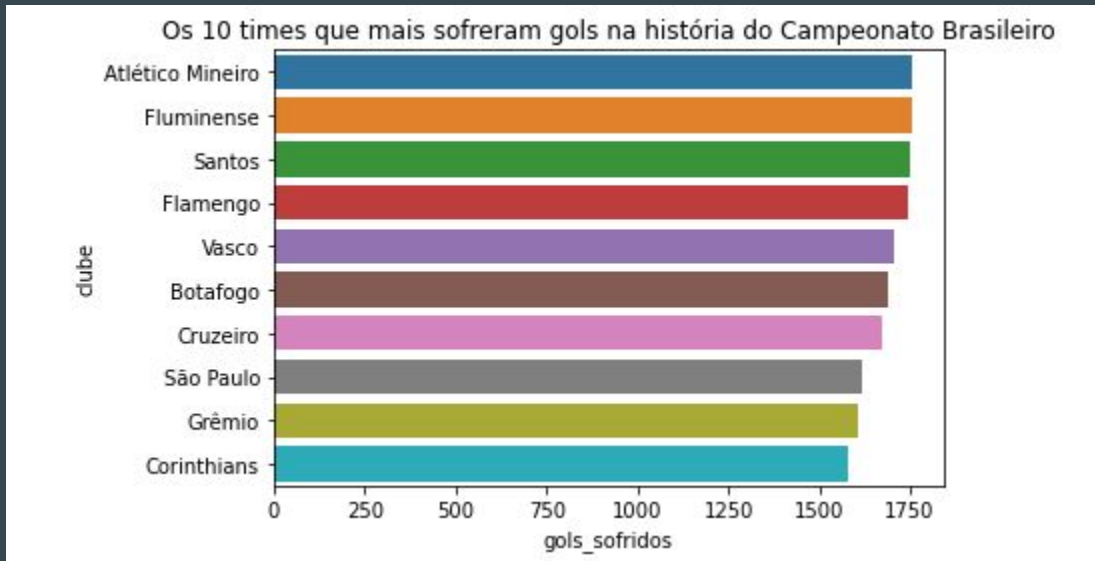
	clube	gols
0	Santos	2274
1	São Paulo	2240
2	Atlético Mineiro	2190
3	Cruzeiro	2145
4	Flamengo	2143
5	Palmeiras	2109
6	Internacional	2049
7	Grêmio	2041
8	Corinthians	1982
9	Vasco	1919

Análises



Análises

O mesmo para os gols sofridos, que também podem ser observados quando um time estava jogando em casa ou fora de casa.



Análises

Diferença na lista de gols sofridos como visitante com relação aos mandantes.

gols_sofridos_visitante

	clube	gols_sofridos
0	Santos	1069
1	Fluminense	1067
2	Flamengo	1040
3	Atlético Mineiro	1040
4	Grêmio	1001

gols_sofridos_mandante

	clube	gols_sofridos
0	Vasco	734
1	Atlético Mineiro	714
2	Flamengo	702
3	Cruzeiro	696
4	Botafogo	690

Análises

Podemos também buscar por todas as vitórias como mandante de um determinado clube, o que tem grande valor historicamente falando. No exemplo a seguir temos o Fortaleza Esporte Clube. (que pudemos ver que venceu 113 dos seus 240 jogos em casa no campeonato brasileiro, 47%).

```
fecvit = df.query('mandante=="Fortaleza" & placar_mandante>placar_visitante')
```

fecvit

	fase	data	horário	mandante	visitante	placar_mandante	placar_visitante
39	1F	1961-07-16	17:00	Fortaleza	ABC	3	1
55	2F	1961-08-20	16:05	Fortaleza	Remo	2	1
56	2F	1961-08-23	21:00	Fortaleza	Remo	2	0
235	3F	1966-08-31	NaN	Fortaleza	Paysandu	3	1
282	3F	1966-08-31	NaN	Fortaleza	Paysandu	3	1
...
20513	R11	2021-07-11	20:30	Fortaleza	Corinthians	1	0
20529	R13	2021-07-25	16:00	Fortaleza	Red Bull Bragantino	1	0
20655	R26	2021-10-13	20:30	Fortaleza	Grêmio	1	0

Análises

A nossa ideia foi trabalhar com uma análise mais focada na parte histórica do torneio, mas por termos todas as partidas em nosso dataset seria possível também fazer análises mais localizadas, como referentes a anos ou períodos específicos de disputa.

FIM