



# Aposta certa

Grupo GoIDB - Trabalho Final MC536  
Segundo Semestre 2021



# Tema do dataset

- Contexto
  - O futebol é o esporte mais popular do mundo
  - O futebol é muito emocionante e, aparentemente, imprevisível
  - O dinheiro é o que move a sociedade contemporânea



# Tema do dataset

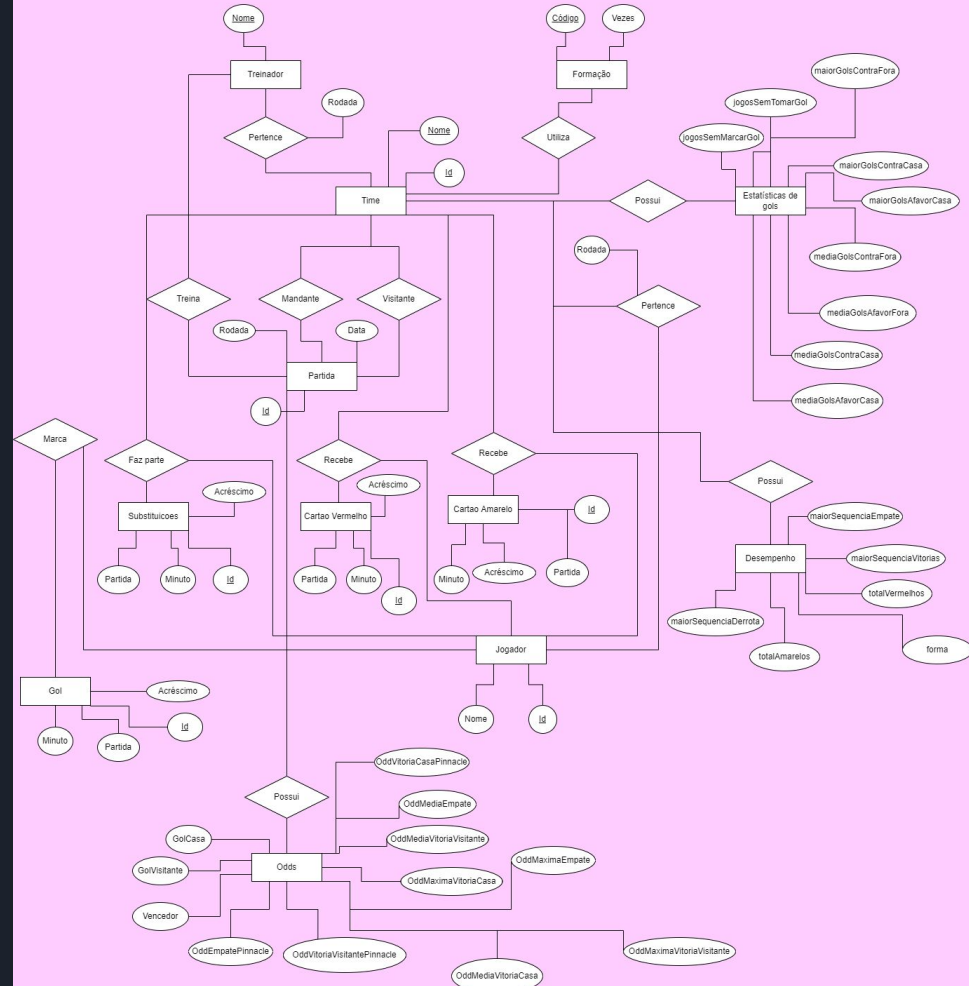
- Contexto
  - O futebol é o esporte mais popular do mundo,
  - O futebol é muito emocionante e, aparentemente, imprevisível
  - O dinheiro é o que move a sociedade contemporânea
- Motivação
  - A produção e o acesso a dados é cada vez mais constante
  - Histórias como a do filme MoneyBall e a de Jimmy Simons permeiam a imaginação e desafiam a ambição dos mais ousados

# Modelo Conceitual

- Entidades

- Treinador
- Time
- Partida
- Gol
- Substituições
- Cartão Amarelo
- Cartão Vermelho
- Jogador

## Modelo Conceitual



# Modelo Lógico Tabular

- Modelo mais estudado em sala de aula

## Modelo Lógico

CARTAOAMARELO(\_Id Cartao\_, Id partida, Id Jogador, Id Time, Acrescimo, Minuto)

Id Jogador chave estrangeira -> JOGADOR(Id jogador)

Id Time chave estrangeira -> TIME(Id Time)

Id partida chave estrangeira -> PARTIDA(Id partida)

CARTAOVERMELHO(\_Id cartao\_, Id partida, Id Jogador, Id Time, Acrescimo, Minuto)

Id Jogador chave estrangeira -> JOGADOR(Id jogador)

Id Time chave estrangeira -> TIME(Id Time)

Id partida chave estrangeira -> PARTIDA(Id partida)

JOGADOR(\_Id jogador\_, Nome, Rodada1, Rodada2, ..., Rodada38)

PARTIDA(\_Id Time\_, Nome Time)

TREINADOR(\_Nome\_, Rodada1, Rodada2, ..., Rodada 38)

GOL(\_Id gol\_, Id partida, Id Time, Id Jogador, Acrescimo, Minuto)

Id Jogador chave estrangeira -> JOGADOR(Id jogador)

Id Time chave estrangeira -> TIME(Id Time)

Id partida chave estrangeira -> PARTIDA(Id partida)

SUBSTITUICOES(\_Id subst\_, Id partida, Id Time, Id Jogador entra, Id Jogador sai, Acrescimo, Minuto)

Id jogador entra chave estrangeira -> JOGADOR(Id jogador)

Id jogador sai chave estrangeira -> JOGADOR(Id jogador)

Id partida chave estrangeira -> PARTIDA(Id partida)

Id Time chave estrangeira -> TIME(Id Time)

FORMACOES(\_Id Time\_, Nome Time, \_Codigo\_, vezes.Jogadas)

Id Time chave estrangeira -> TIME(Id Time)

CLUBESDESEMPENHO(\_Id Time\_, Nome time, maiorSequenciaEmpate, maiorSequenciaDerrota, maiorSequenciaVictorias, totalAmarelos)

Id Time chave estrangeira -> TIME(Id Time)

CLUBES\_GOLSTATS(\_Id Time\_, Nome Time, jogosSemMarcarGol, jogosSemTomarGol, maiorGolsContraFora, maiorGolsContraCasa, r

Id Time chave estrangeira -> TIME(Id Time)

# Modelo Lógico de Documentos


- Modelo condizente com a estrutura hierárquica dos times e do campeonato

Modelo Lógico (JSON)



# Fontes de dados

- API-Football:
  - Fixtures (partidas)
  - Teams
  - Events
  - Lineups



Q Search...

- Introduction
- Authentication >
- Architecture
- Sample Scripts >
- WIDGETS
  - Widgets >
- ENDPOINTS
  - Timezone >
  - Countries >
  - Leagues >
  - Teams >
  - Venues >

## API-FOOTBALL (3.8.1)

support: <https://dashboard.api-football.com> | URL: <https://www.api-football.com>

### Introduction

Welcome to Api-Football! You can use our API to access all API endpoints, which can get information about Football Leagues & Cups.

We have language bindings in C, C#, cURL, Dart, Go, Java, Javascript, Node.js, Objective-c, OCaml, Php, PowerShell, Python, Ruby, Shell and Swift! You can view code examples in the dark area to the right, and you can switch the programming language of the examples with the tabs in the top right.

We have a **Public Roadmap** on Github, which will allow you to follow all the upcoming developments but also the possibility for everyone to suggest new features.

To propose a new feature the procedure is described [here](#).

The Roadmap is available via this link: <https://github.com/api-sports/public-roadmap/projects/1>.

This roadmap is only the improvements and new features and not the bugs, for the most part we manage them directly as we see them or as they are reported by our users. To report a bug, you can do it directly from the chat on the Dashboard.

### Authentication

# Fontes de dados

- Football-Data.co.uk:
  - Resultados das probabilidades de apostas de encerramento (casa-empate-visitante) da Pinnacle Sports, bem como os odds máximos e média dentre uma série de casas de apostas. Dados do campeonato brasileiro de 2012 a 2021

**Football-Data.co.uk**  
Results / Odds / Tipsters

Network Sites ▼ Updated: 22/11/21 SHARE BeGambleAware Follow

Home Free Bets Livescores Books on Betting Casino Poker Tennis Contact Like

**WORLD'S FAVOURITE**

**TOP RATED**

- BoyleSports
- William Hill
- Paddy Power
- Betfred
- Betway

**NEW CUSTOMERS**

- bet365 Bet Credits
- 630 Free Bets

**Data Files: Brazil**  
Last updated: 19/11/21

Registering with any of the advertised bookmakers on [Football-Data](#) will help keep access to the historical results & betting odds data files FREE.

Below you will find download links for full time results and closing match (home-draw-away) betting odds from as well as best and average market prices from a basket of bookmakers. Data are available in both Excel and CSV format and are available from 2012 onwards.

Notes.txt  
(text file key to the data files and data source acknowledgements)

Excel CSV

Contact [Football-Data.co.uk](#) if you believe there are any errors or omissions in the data files.

**SITE RESOURCES**

- Historical Data
- Learn to Bet
- Free Bets
- Books
- Other Sites

**BETTING SYSTEMS**

- Football Ratings
- Wisdom of Crowds
- Contrarian Betting

**BETTING ARTICLES**

- Football-Data
- Pinnacle Sportsbook





# Extração e Tratamento

- Requisições à API via Python Requests (GET):

```
url = "https://v3.football.api-sports.io/{endpoint}"

payload={}
headers = {
    'x-rapidapi-key': 'XxXxXxXxXxXxXxXxXxXxXxXx',
    'x-rapidapi-host': 'v3.football.api-sports.io'
}

response = requests.request("GET", url, headers=headers, data=payload)
```



## Extração e Tratamento - Intermediários

- fixtures.json
- fixtures-events.json
- fixtures-lineups.json
- estatisticasDeDesempenho.pkl
- estatisticasDeGol.pkl
- formacoesUsadas.pkl

# Extração e Tratamento

- Leitura direta do CSV de Odds usando Pandas

Country	League	Season	Date	Time	Home	Away	HG	AG	Res	PH	PD	PA	MaxH	MaxD	MaxA	AvgH	AvgD	AvgA
Brazil	Serie A	2012	19/05/2012	22:30	Palmeiras	Portuguesa	1	1	D	1.75	3.86	5.25	1.76	3.87	5.31	1.69	3.5	4.9
Brazil	Serie A	2012	19/05/2012	22:30	Sport Recife	Flamengo RJ	1	1	D	2.83	3.39	2.68	2.83	3.42	2.7	2.59	3.23	2.58
Brazil	Serie A	2012	20/05/2012	01:00	Figueirense	Nautico	2	1	H	1.6	4.04	6.72	1.67	4.05	7.22	1.59	3.67	5.64
Brazil	Serie A	2012	20/05/2012	20:00	Botafogo RJ	Sao Paulo	4	2	H	2.49	3.35	3.15	2.49	3.39	3.15	2.35	3.26	2.84
Brazil	Serie A	2012	20/05/2012	20:00	Corinthians	Fluminense	0	1	A	1.96	3.53	4.41	1.96	3.53	4.41	1.89	3.33	3.89
Brazil	Serie A	2012	20/05/2012	20:00	Internacional	Coritiba	2	0	H	1.76	3.63	5.66	1.77	3.9	6	1.66	3.55	5.1
Brazil	Serie A	2012	20/05/2012	20:00	Ponte Preta	Atletico-MG	0	1	A	3.12	3.45	2.44	3.12	3.45	2.62	2.74	3.23	2.45



# Queries e Manipulação de Dados

- Banco de Dados está sendo implementado em SQLite.
- Através da leitura dos arquivos intermediários, lemos e formatamos arquivos .csv de cada tabela:
  - cartoes\_amarelos.csv
  - gols.csv
  - partidas.csv
  - treinadores.csv
  - cartoes\_vermelhos.csv
  - jogadores.csv
  - substituicoes.csv
  - formacoes.csv
  - times.csv
- Esses arquivos são transformados em tabelas no nosso Banco SQLite, em que é possível realizar operações na linguagem SQL.

# Tabelas formadas - Exemplos

Home	Away	HG	AG	Res	PH	PD	PA	MaxH	MaxD	MaxA	AvgH	AvgD	AvgA	PartidaID
CSA	Palmeiras	1.0	1.0	D	7.04	3.93	1.56	7.80	4.00	1.56	6.88	3.85	1.51	94485
Avai	Gremio	1.0	1.0	D	3.88	3.16	2.18	4.20	3.30	2.21	3.72	3.11	2.10	94481
Vasco DA Gama	Corinthians	1.0	1.0	D	3.26	3.16	2.46	3.26	3.20	2.50	3.10	3.06	2.39	94500
Chapecoense-sc	Atletico Paranaense	1.0	1.0	D	2.64	2.99	3.15	2.66	3.13	3.20	2.49	2.96	3.04	94494
CSA	Santos	0.0	0.0	D	4.84	3.49	1.85	5.00	3.55	1.89	4.57	3.37	1.81	94496
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

idGol	idPartida	idTime	idJogadorGol	idJogadorAssistencia	tempo	acrescimo
1	94471	134	10135	10150.0	3	
2	94471	134	10150		45	
3	94471	134	10550		66	
4	94471	134	10133	10143.0	86	
5	94471	133	10573	10574.0	89	
6	94472	1062	9876		45	2.0
7	94472	145	10538	10518.0	46	
8	94472	1062	9894	9901.0	53	
9	94473	131	10244	10249.0	45	2.0

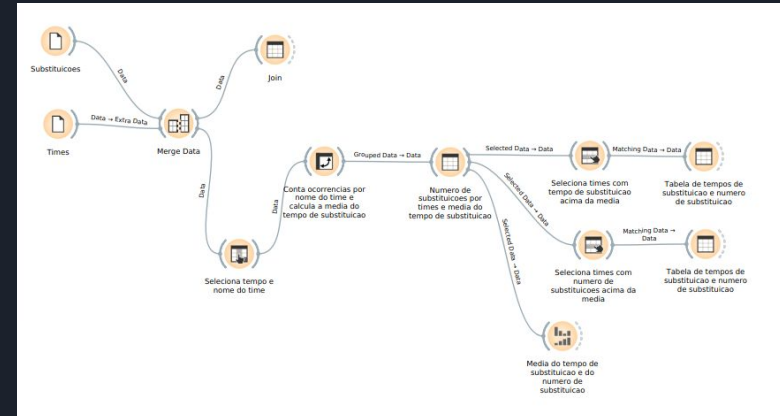
# Perguntas de pesquisa

1. As substituições realizadas durante uma partida se relacionam com o desempenho do time no campeonato?

## Etapas realizadas

- Dar join nas tabelas de time e de substituição
- Selecionar apenas as colunas de tempo de substituição e número de substituições
- Calcular a média de tempo de substituição total e a média de número de substituições
- Selecionar os times com estatísticas acima da média
- Verificar o desempenho do time no campeonato

Resultado: Times que tem um bom desempenho geralmente realizam menos substituições durante uma partida.





# Perguntas de pesquisa

2. Se fizermos apostas sempre de acordo com os Odds médios, qual a chance de acerto?

Etapas realizadas:

- Da tabela de partidas, selecionar todas as colunas com AvgA > AvgH e Res = H
- Da tabela de partidas, selecionar todas as colunas com AvgH > AvgA e Res = A
- Dar join em ambas as seleções
- Contar quantas linhas

Resultado: cerca de **53%** de chance de acerto

```
cur = con.cursor()
cmd = '''CREATE VIEW R1 AS SELECT * FROM Partidas P where P.AvgA > P.AvgH AND P.Res = 'H';'
cur.execute(cmd)
```

```
ur = con.cursor()
cmd = '''CREATE VIEW R2 AS SELECT * FROM Partidas P where P.AvgH > P.AvgA AND P.Res = 'A';'
cur.execute(cmd)
```



# Perguntas de pesquisa

3. Defesas mais sólidas tem impacto nas vitórias fora de casa?
  - . Criamos duas views: Resultados de Partidas com vitórias de time visitante e Estatísticas de gols
  - . Ordenamos as duas views e analisamos as posições de cada time nas duas tabelas





# Perguntas de pesquisa não respondidas

## 1. Jogadores que marcam muitos gols são escalados para mais partidas ?

- Agrupar os registros da tabela de gols pelo nome do jogador e contar ocorrências
- Calcular a média de partidas jogadas pelos jogadores que marcam gols
- Calcular a média de partidas jogadas no geral
- Comparar os resultados



# Perguntas de pesquisa não respondidas

## 2. Jogadores que receberam cartão numa partida são frequentemente substituídos ?

- Selecionar na tabela de cartões amarelos Time mandante, Time visitante e nome do jogador
- Selecionar na tabela de substituições Time mandante, Time visitante e nome do jogador de jogadores que façam parte da seleção anterior
- Contar o número de registros
- Selecionar na tabela de substituições Time mandante, Time visitante e nome do jogador de jogadores
- Contar o número de registros
- Comparar os resultados



# Perguntas de pesquisa não respondidas

3. Marcar gols no começo da partida geralmente resulta em vitória ou os times se recuperam facilmente no decorrer da partida ?

- Agrupar os registros da tabela de gols pelas partidas e selecionar os tempos mínimos
- Verificar se o time marcador foi o time que ganhou o jogo.