

Metrificação e análise de carreiras de jogadores de futebol por meio de um sistema computacional

Lorenzo Bertin Salvador

Orientadora: Prof^a. Dr^a. Kelly Rosa Braghetto

Contexto e Motivação

- Futebol é o esporte mais popular do mundo.
- Discussões incessantes para decidir os melhores jogadores
 - De cada ano
 - De cada time
 - Da história



- As premiações mais prestigiadas não têm critérios claros, gerando incoerências ao longo dos anos.
- Pergunta: É possível avaliar a qualidade de um jogador por critérios sistemáticos?

É possível avaliar a qualidade de um jogador por critérios sistemáticos?

SIM, de acordo com o artigo: *Modk: formula for determining the best season and career of a football player by objective indicators (Salvador et al, 2022)*.

- Produz uma pontuação, baseada em uma fórmula matemática, para toda a carreira de um jogador.
- Abrange tanto jogadores atuais quanto quem já se aposentou.
- Leva em conta informações sobre os times, campeonatos, países e continentes.

Objetivos

1. Sistema capaz de implementar a fórmula;
2. Análise comparativa entre resultados da fórmula e outra maneira de se mensurar a performance de um jogador;
3. Possibilidade de usuários com pouco conhecimento de computação interagirem com o sistema.

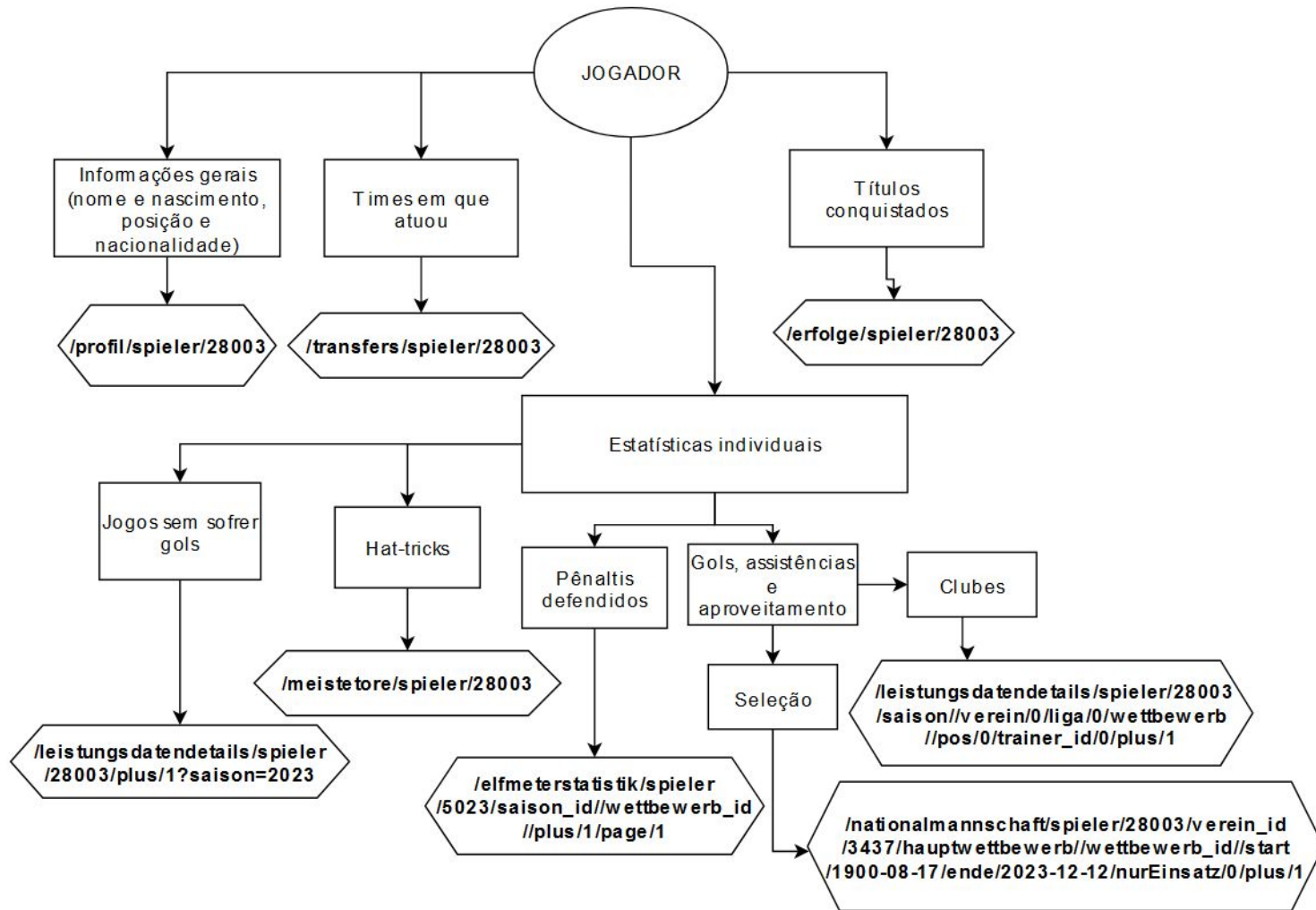
Organização do Sistema

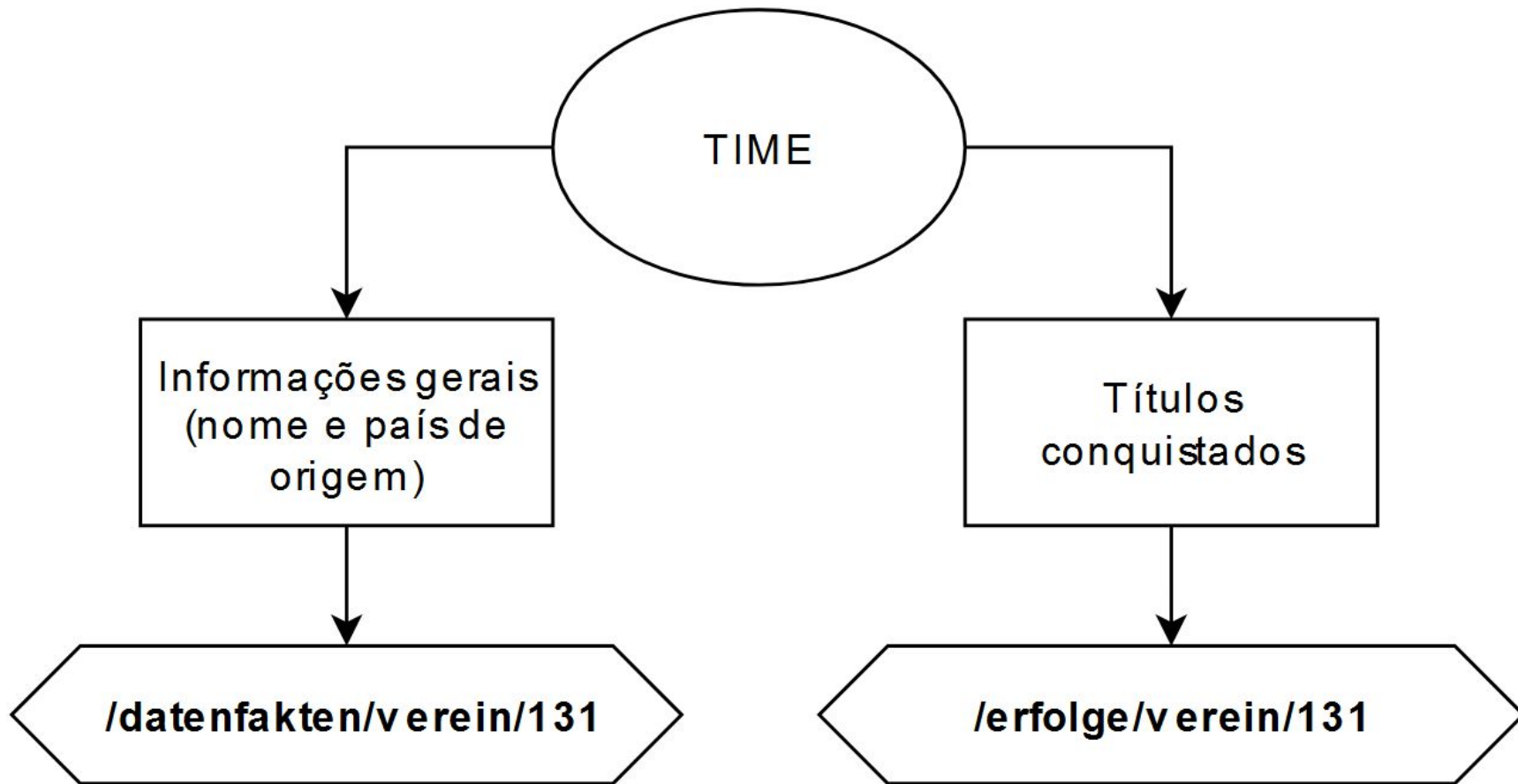
- **Componentes centrais:** essenciais para aplicação da fórmula
 - Extrator de dados da Internet
 - Banco de dados operacional
 - Calculadora da fórmula
 - Interface de gerenciamento e visualização dos dados
- **Componente auxiliar:** enriquece a análise fornecendo perspectiva diferente
 - Banco de dados de grafos

Extração de Dados

- Fonte: **Transfermarkt**
- Plataforma não possui nenhum serviço de API.
- Solução: *Web Scraping*.
- Informações obtidas:
 - Times:
 - Nome e país de origem
 - Títulos
 - Jogadores:
 - Nome e informações gerais
 - Estatísticas individuais
 - Títulos conquistados
 - Times em que atuou
- Também é possível complementar os dados manualmente (planilhas CSV).

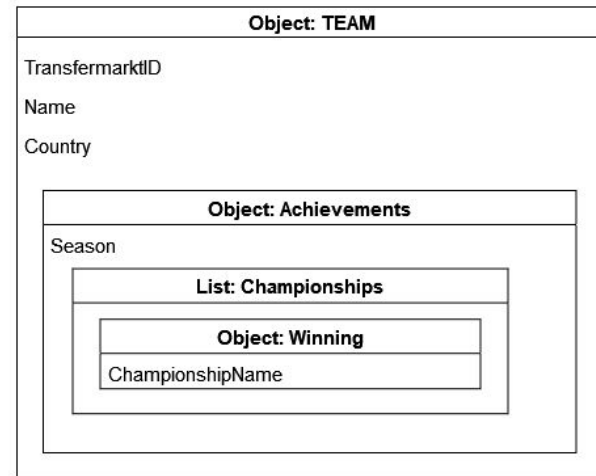
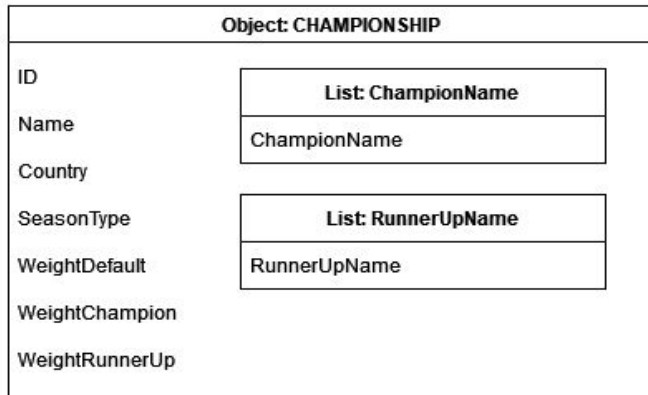


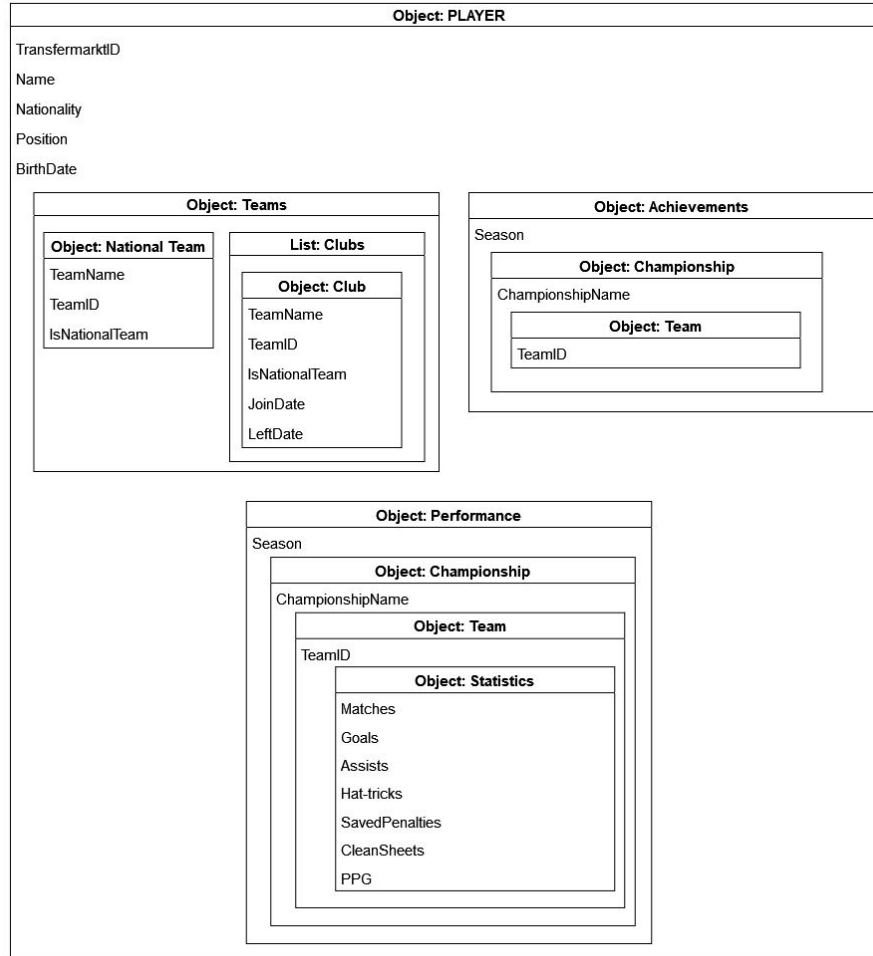




Banco de Dados - Operacional

- Orientação a documentos: **MongoDB**.
- Justificativa: documento é forma natural de se armazenar os dados advindos da extração.
- Coleções: **PLAYER, TEAM, SCORE, CHAMPIONSHIP, COUNTRY e CONTINENT.**





Object: Performance

Season

Object: Championship

ChampionshipName

Object: Team

TeamID

Object: Statistics

Matches
 Goals
 Assists
 Hat-tricks
 SavedPenalties
 CleanSheets
 PPG

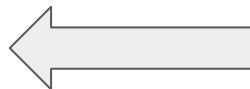
Banco de Dados - Grafo

- SGBD: **Neo4j**.
- Vértices: Jogadores.
- Arestas: Atuação conjunta de dois jogadores em certo campeonato.
- Trata o conjunto de jogadores coletivamente.
- Estrutura do grafo fornece métricas próprias:
 - Maior menor caminho
 - Componentes conexas
 - Quantidade de triângulos
 - K-1 cores
 - Grau de centralidade
 - Máximo de influência
- Comparação entre pontuações de um jogador com seus vizinhos.

Lionel Messi	
Pontuação total:	8491
Pontuação individual:	5115
Pontuação (média) total vizinhos:	2475
Pontuação (média) individual vizinhos:	847
Total de vizinhos:	46

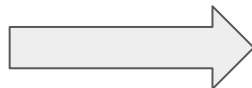
Exemplo de métricas obtidas no Neo4J

```
CALL gds.allShortestPaths.stream('graph')
YIELD sourceNodeId, targetNodeId, distance
WITH sourceNodeId, targetNodeId, distance
WHERE isFinite(distance) = true
WITH asNode(sourceNodeId) AS source,
asNode(targetNodeId) AS target, distance
WHERE source <> target
RETURN source, target, distance
```



Maior menor caminho

Grau de centralidade



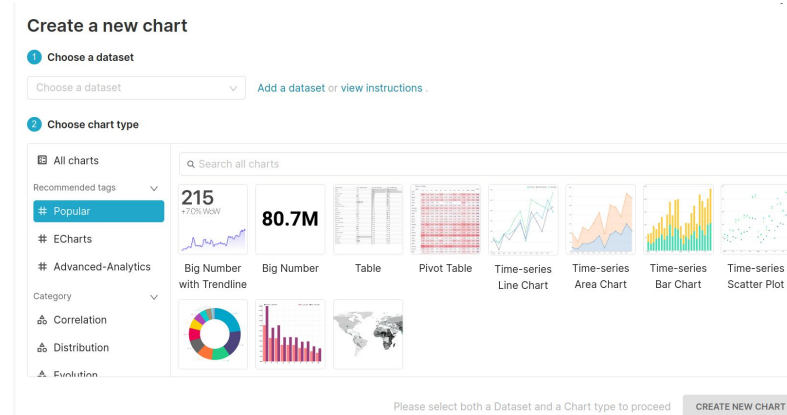
```
CALL gds.degree.stream('graph')
YIELD nodeId, score
RETURN gds.util.asNode(nodeId).name
AS name, score AS followers
ORDER BY followers DESC, name DESC
```

Aplicação da Fórmula

- **Pontuação individual:** avalia as estatísticas individuais do jogador.
 - Atributos com certo peso: gols, assistências, hat-tricks, jogos sem sofrer gols, pênaltis defendidos, aproveitamento.
 - Soma número de ocorrências * peso de cada atributo.
- **Pontuação coletiva:** avalia os títulos conquistados pelo jogador.
 - Soma-se o peso de cada um dos campeonatos presentes na base.
- **Pontuação de impacto:** avalia o impacto causado em cada time que atuou
 - Impacto agudo: impacto imediato da entrada do jogador em certo time.
 - Impacto de longo prazo: pernilagem de participação de um jogador em todos os títulos de um time.
- **Coeficientes de ajuste:** ajusta cada campeonato/título à sua dificuldade e contexto
 - Pesos para campeonatos, países, continentes e temporadas.
 - Ganhar um campeonato continental na Europa é mais valioso que um continental na Ásia.
 - Ganhar a Copa do Mundo é mais valioso que o Campeonato Paulista.

Interface de Usuário

- **Mongo Compass:** inserção dos dados estáticos (países, continentes e campeonatos)
- **Interface customizada:** inserção dos dados dinâmicos (jogadores e times)
- **Neo4j:** Gerenciamento e visualização do grafo
- **Apache Superset:** visualização e manipulação das pontuações dos jogadores



```
neo4j$ MATCH (n) RETURN n LIMIT 300
```



Graph



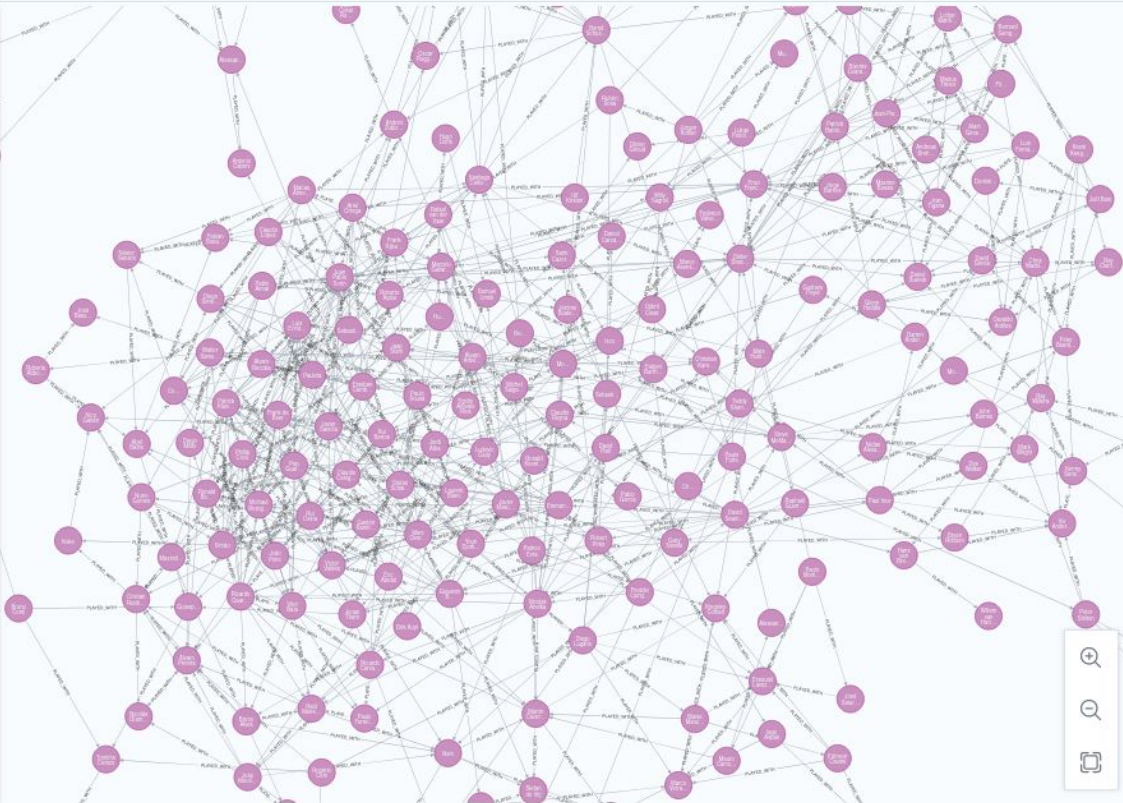
Table



Text



Code



Overview

Node labels

* (300)

Player (300)

Relationship types

* (880)

PLAYED_WITH (880)

Displaying 300 nodes, 0 relationships.

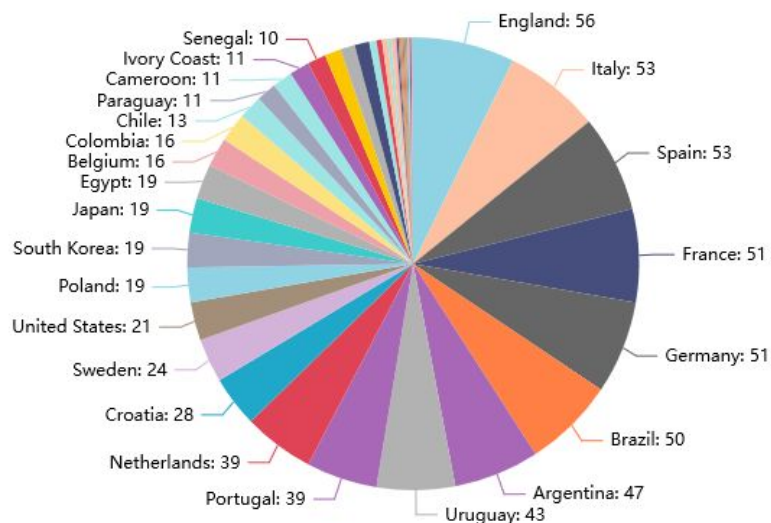


Resultados: Demografia dos jogadores

Players countries distribution - PIE

England Italy Spain France Germany Brazil Argentina Uruguay 1/7

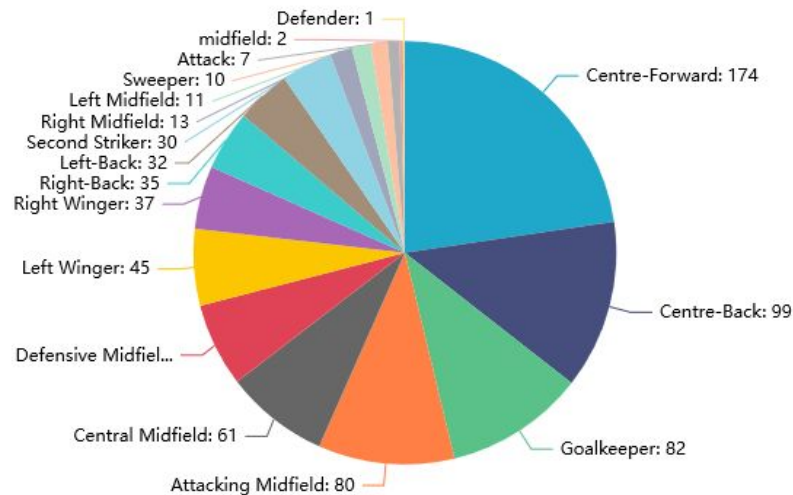
Total: 768



Players positions

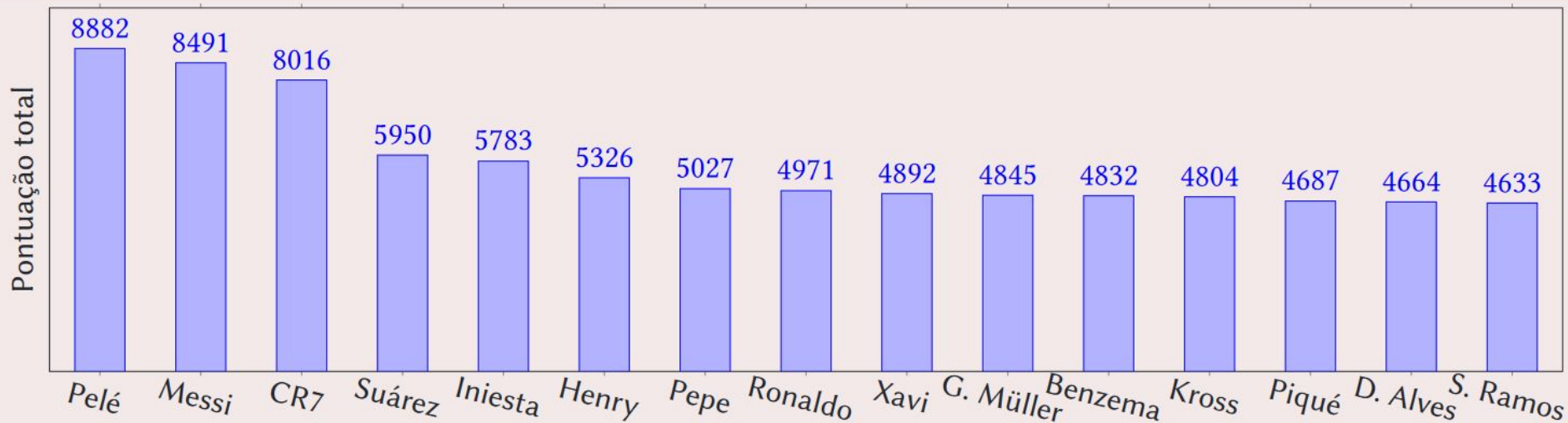
Centre-Forward Centre-Back Goalkeeper Attacking Midfield 1/5

Total: 768



Resultados: Top 15

Pontuações totais - Top 15



Resultados: Comparação Fórmula - France Football

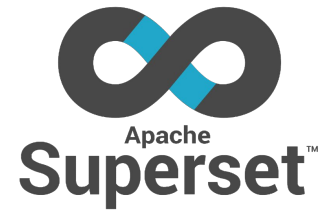
Ano	Vencedor fórmula	Vencedor <i>France Football</i>
2023	Lionel Messi	Lionel Messi
2022	Jorginho	Karim Benzema
2021	Riyad Mahrez	Lionel Messi
2020	Mohamed Salah	-
2019	Kylian Mbappé	Lionel Messi
2018	Cristiano Ronaldo	Luka Modric
2017	Cristiano Ronaldo	Cristiano Ronaldo
2016	Claudio Bravo	Cristiano Ronaldo
2015	Toni Kroos	Lionel Messi
2014	Cristiano Ronaldo	Cristiano Ronaldo
2013	Cesc Fàbregas	Cristiano Ronaldo
2012	Lionel Messi	Lionel Messi
2011	David Villa	Lionel Messi
2010	Éric Abidal	Lionel Messi
2009	Xavi	Lionel Messi
2008	Carlos Tevez	Cristiano Ronaldo
2007	Marco Materazzi	Kaká
2006	Ronaldinho	Fabio Cannavaro
2005	Deco	Ronaldinho
2004	Thierry Henry	Andriy Shevchenko
2003	Ronaldo	Pavel Nedvěd
2002	Víctor Aristizábal	Ronaldo
2001	Bixente Lizarazu	Michael Owen
2000	Pavel Nedved	Luís Figo

Conclusão

- Base para a produção de mais análises futebolísticas, potencialmente realizadas por agentes de outras áreas do conhecimento
- Mesmo sem a fórmula, a estruturação e armazenamento dos dados é valiosa
- Possibilidades de evolução:
 - Mais inserções de jogadores, times e campeonatos
 - Mais fontes além do *Transfermarkt*

Principais tecnologias

- BeautifulSoup
- MongoDB
- Neo4j
- Apache Superset
- Flask



Repositório: <https://gitlab.com/lorenzobs/sistema-modok-publico>

Referências para esta apresentação:

<https://www.footyrenders.com/render/lionel-messi-ballon-dor-2023.png>

https://upload.wikimedia.org/wikipedia/commons/thumb/7/7b/Transfermarkt_logo.png/800px-Transfermarkt_logo.png

<https://repository-images.githubusercontent.com/39464018/58649580-eca4-11ea-844d-c2ddca24b226>

https://upload.wikimedia.org/wikipedia/commons/thumb/9/93/MongoDB_Logo.svg/1280px-MongoDB_Logo.svg.png

https://upload.wikimedia.org/wikipedia/commons/e/e5/Neo4j-logo_color.png

https://upload.wikimedia.org/wikipedia/commons/thumb/3/3c/Flask_logo.svg/1280px-Flask_logo.svg.png

https://miro.medium.com/v2/resize:fit:772/0*oN9jA-Ad3mRlPAYy.png