

Módulo VI: Mestrado BIG DATA APLICADO AO FUTEBOL

EXERCÍCIO PRÁTICO

Aluno: Cassio Félix Jardim
email: cassiofjardim@gmail.com
(<mailto:cassiofjardim@gmail.com>)

Escolhendo um Provedor de dados

Você foi contratado para gerir o departamento de Analytics de um grande clube Português. Foi-lhe pedido para contratar um fornecedor de dados desportivos para começar o projeto. Qual seria a empresa que você escolheria e Porque?

Provedores Pagos



Uma das maiores empresas de Base de Dados no Futebol. O fornecimento dos dados é pago e de alto valor. Os destaques principais são os dados não-estruturados (vídeos e imagens por exemplo)

STATSBOMB

Empresa que investe na análise de dados estruturados e não estruturados (vídeos e imagens, por exemplo). Além disso, a empresa é muito engajada em estar trazendo novidades sobre novas métricas na análise de futebol. Vale muito a pena acompanhar as atualizações em sua conta do [Twitter](https://twitter.com/statsbomb) (<https://twitter.com/statsbomb>) (<https://twitter.com/statsbomb>). Dependendo do tipo de dado o acesso é livre.

wyscout

Outra grande fornecedora de dados para o futebol. A WYSCOUT possui planos para o acesso aos dados, sendo acessível a clubes, agentes e pesquisadores. [wyscout](https://wyscout.com/pricing/) (<https://wyscout.com/pricing/>)

Provedores Livres



FBREF é uma base de dados muito interessante por ser vasta e de acesso livre. Dependendo da liga de interesse (Premiere League, La Liga, Serie A, por exemplo), pode-se fazer um estudo detalhado de equipes e jogadores utilizando seus dados.



Transfermarkt é uma base de dados sobre finanças do futebol. Base de dados com informações pessoais e de valor de mercado de jogadores e clubes.



Understat segue na linha da FBREF. No entanto, a UNDERSTAT fornece algumas métricas que não estão presentes na FBREF.



Para termos acesso de uma forma facilitada e direta as bases de dados da FBREF, Footmob, Transfermarkt e Understat, podemos utilizar o pacote `worldfootballR` para a linguagem R. Esse é o link para o repositório do pacote: [worldfootballR package](https://github.com/JaseZiv/worldfootballR) (<https://github.com/JaseZiv/worldfootballR>)

Conclusão:

Dessa forma podemos concluir que as empresas que fornecem dados sob uma assinatura paga estão na fronteira da criação, extração e fornecimento dos dados para o Futebol. No entanto, utilizar de bases de dados como FBREF e Understat, e mais ainda, cruzar essas bases de dados pode oferecer insights valiosos sobre jogadores e equipes.

Data Scouting

A primeira aplicação do departamento seria através do Scouting. A equipe precisa de um avançado para substituir outro que acaba de sair. Selecione 4 métricas que você utilizaria para identificar o substituto e explique os motivos.

≡ Average Expected Goals (xG) vs Average goals:

Gols esperados (xG): É uma das métricas utilizadas hoje. Para atacantes xG traz uma importante referência sobre a eficácia/eficiência dos artilheiros. Além disso, relacionar essa métrica com a média de gols na temporada reforça a qualidade dessa métrica. Quanto maior for a média de gols diante do xG do atleta mais eficiente esse atacante será.

≡ Average xG vs Average Touches in the box:

Novamente iremos relacionar duas métricas. A media de gols Esperados e o numero de toques na bola dentro da grande area. Dessa forma, iremos ter alguma indicação da qualidade das ações desse atacante dentro da grande area. Quanto menor for a quantidade de toques dentro da grande area relativa a media de gols esperados mais eficiente será esse atacante.

≡ Average Shots vs Average Shots on target:

Relacionar o numero de Chutes com o numero de chutes no alvo, reflete a precisão das finalizações desse atleta. Quanto maior for a razão $\frac{\text{Average Shots on Target (SoT)}}{\text{Average Shots (Sh)}}$ maior o nivel de precisão desse atacante.








≡ Average Dribbles vs Average Shots:

Estabelecer uma relação entre uma media de dribles e uma media de chutes pode ser um bom indicador da característica do jogador. Se estivermos pensando em um Atacante mais fixo de Area talvez seja bem provável que a media de chutes seja superior seja que a media de dribles. Para um jogador de lado de campo é esperado que o oposto aconteça.

Transfermarkt

O seu presidente definiu que é necessário que o substituto seja um jogador livre no mercado e sub23. Vá até o Transfermarkt e faça uma lista de jogadores que se encaixam no perfil desejado.


Top Atacantes Sub-23 Livres no Mercado

COMP.	EQUIPE	JOGADOR	IDADE	PAIS	ALTURA (M)	DATA CONTRATO
Liga Bwin	FC Arouca	Antony	20		1.78	30/06/2022
Premier League	Southampton FC	Armando Broja	20		1.91	31/05/2022
Serie A	Spezia Calcio	Eddie Salcedo	20		1.78	29/06/2022
Ligue 1	Paris Saint-Germain	Kylian Mbappe	23		1.78	30/06/2022
Ligue 1	Stade Brestois 29	Martin Satriano	20		1.87	30/06/2022
Liga Bwin	Portimonense SC	Renato Junior	19		1.9	30/06/2022
Liga Bwin	CS Marítimo	Ricardinho	20		1.81	30/06/2022

Source: [Transfermarkt](https://fbref.com/en) (https://fbref.com/en)
Temporada 2021/2022


Comentários

Após filtragem dos dados da Transfermarkt chegamos a uma lista de 7 jogadores que atendem as condições exigidas: serem sub-23 e com contrato expirando ou sem contrato. Todos os jogadores fazem parte das principais ligas da Europa. São elas: Liga Portuguesa(Bwin), Liga Francesa, Liga A italiana e Premier League inglesa. A princípio, o fato de serem jogadores de liga de grande expressão na Europa conta como um ponto positivo na hora da escolha dos atletas. Mas podemos fazer mais. Analisando outras informações no Transfermarkt podemos estabelecer um grau de prioridade na escolha dos atletas. Abaixo apresentamos análises individuais sobre os atletas levando em conta informações como: *tempo fora por causa de lesões, quantidade de jogos perdidos por causa de lesões, Gols, numero de assistencias(Ast.) e gols esperados (xG)* (informações valiosas para um bom atacante):

Antony 	
INFO.	VALUE
Dias Fora (Lesão)	0.0
Jogos Fora (Lesão)	0.0
Min	402.0
Gls	0.0
Ast	0.0
xG	1.4
Source: FBREF website (https://fbref.com/en) Temporada 2021/2022	

Prós: Jogador com pouco histórico de contusões.

Contra: Pouca minutagem na temporada.

Armando Broja 	
INFO.	VALUE
Dias Fora (Lesão)	27.0
Jogos Fora (Lesão)	5.0
Min	1862.0
Gls	12.0
Ast	10.0
xG	13.4
Source: FBREF website (https://fbref.com/en) Temporada 2021/2022	

Prós: Jogador com boa media de gols, apesar das lesões que o prejudicaram na temporada. Um bom nivel de Assistências e Xg. Boa minutagem. Uma boa aposta.

Contra: As lesões impactaram no seu aproveitamento. A altura de 1.9m indica uma limitação na posição que ele pode ser utilizado.

Transfermarkt(Cont.)

Eddie Salcedo



INFO.	VALUE
Dias Fora (Lesão)	50.0
Jogos Fora (Lesão)	10.0
Min	1146.0
Gls	6.0
Ast	0.0
xG	5.4

Source: [FBREF website](https://fbref.com/en)
(<https://fbref.com/en>)
Temporada 2021/2022

Prós: Jogador com minutagem boa. Números de Gols superior ao xG, o que é um bom indicador para um atacante.

Contra: Um dos jogadores que mais contusões sofreu. O jogador está na média dos seus concorrentes. Vale observar mais seus jogos.

Kylian Mbappe



INFO.	VALUE
Dias Fora (Lesão)	66.0
Jogos Fora (Lesão)	9.0
Min	95.0
Gls	2.0
Ast	0.0
xG	1.1

Source: [FBREF website](https://fbref.com/en)
(<https://fbref.com/en>)
Temporada 2021/2022

Prós: Jogador acima da media internacional. Uma das principais estrelas do campeonato Francês. Seus números mostram isso. Numero de gols acima de xG. Jogador que atua bem nos lados do campo e por dentro.

Contra: Poucos fatores pesam contra o jogador.

Martin
Satriano



INFO.	VALUE
Dias Fora (Lesão)	12
Jogos Fora (Lesão)	2
Min	217
Gls	1
Ast	0
xG	NA

Source: [FBREF website](https://fbref.com/en)
(<https://fbref.com/en>)
Temporada 2021/2022

Prós: De todos os jogadores foi o que menos jogou. Pouca informações sobre o atleta

Contra: De todos os jogadores foi o que menos jogou. Pouca informações sobre o atleta

Renato Junior



INFO.	VALUE
Dias Fora (Lesão)	0
Jogos Fora (Lesão)	0
Min	341
Gls	0
Ast	0
xG	NA

Source: [FBREF website](https://fbref.com/en)
(<https://fbref.com/en>)
Temporada 2021/2022

Prós: Jogador mais jovem da lista. Vale observar melhor seu rendimento nos próximos jogos.

Contra: Poucos jogos. Pouca informações sobre o atleta.

Conclusão:

Após realizarmos um primeiro e segundo filtro, chegamos a conclusão que dos 7 jogadores poderíamos estabelecer 3 prioridades:



Kylian Mbappe:

O jogador mais valioso da lista. Vale o investimento porque o atleta encontra-se no auge de sua forma e com potencial de crescimento.



Armando Broja:

Jogador que possui bons números mesmo com as lesões na temporada. Além disso, atua na principal liga da Europa o que pesa muito ao seu favor.



Eddie Salcedo:

Uma ótima aposta. Tem bons números no campeonato italiano.