



footure
academy

ESPECIALIZAÇÃO

CIÊNCIA NOVO
DE DADOS

NO FUTEBOL

Aula 16 - Linguagens de programação

Caio Batatinha

AGENDA

- Definição dos conceitos
- Exemplos de bibliotecas essenciais
- Exemplos de bibliotecas focadas no futebol

OBJETIVOS DE APRENDIZAGEM

- Aprender o que é uma biblioteca
- Visualização geral das bibliotecas e pacotes chaves para análise de dados no futebol

BIBLIOTECAS - DEFINIÇÃO

Bibliotecas são um conjunto de funções e módulos pré definidos para performar tarefas mais específicas. Seu objetivo é facilitar a lógica de programação com soluções de utilização imediata.

Exemplo: Numpy (manipulação numérica) & ggplot (visualização de dados)

PACOTES - DEFINIÇÃO

De forma análoga, pacotes são um conjunto de bibliotecas que englobam um kit pré pronto para auxiliar no desenvolvimento de código de uma tarefa. Já inclui normalmente uma documentação para auxiliar na operação da biblioteca

Exemplo: Pandas (manipulação de dados) & scikit learn (machine learning)

LINGUAGENS DE PROGRAMAÇÃO



**BIBLIOTECA
S
ESSENCIAIS**

PANDAS

O Pandas é essencialmente um pacote para manipulação de dados. Ele é essencial para todos que trabalham com data science e permite ler diferentes arquivos e transformar num data frame por exemplo, recebendo arquivos como csv,xml,xlsx,json,paquet. Além disso já contém funções predefinidas para criar novas colunas, agrupamento de linhas e colunas, pivotar tabela, fusão de outras tabelas, exclusão ou substituição de valores não existentes.

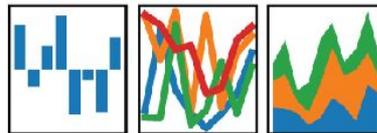
Paralelamente o que se utiliza em R: dplyr e tidyverse

link para documentação:

<https://pandas.pydata.org/docs/index.html>

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



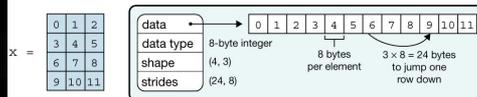
	BandName	WavelengthMax	WavelengthMin
0	CoastalAerosol	450	430
1	Blue	510	450
2	Green	590	530
3	Red	670	640
4	NearInfrared	880	850
5	ShortWaveInfrared_1	1650	1570
6	ShortWaveInfrared_2	2290	2110
7	Cirrus	1380	1360

NUMPY

Biblioteca importante para manipulação numérica.
Essencial na transformação de dados principalmente ao lidar com diferentes formatos como: string, float, int, array.
Utilizado para tarefas de álgebra linear e operações estatísticas
Paralelamente o que se utiliza em R: matrixStats

link para documentação:
<https://numpy.org/>

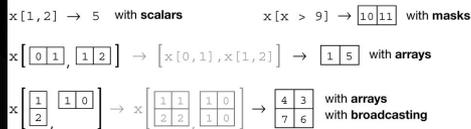
a Data structure



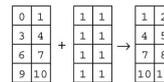
b Indexing (view)



c Indexing (copy)



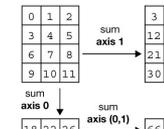
d Vectorization



e Broadcasting



f Reduction



g Example

```
In [1]: import numpy as np
In [2]: x = np.arange(12)
In [3]: x = x.reshape(4, 3)

In [4]: x
Out [4]:
array([[ 0,  1,  2],
       [ 3,  4,  5],
       [ 6,  7,  8],
       [ 9, 10, 11]])

In [5]: np.mean(x, axis=0)
Out [5]: array([4.5, 5.5, 6.5])

In [6]: x = x - np.mean(x, axis=0)

In [7]: x
Out [7]:
array([[ -4.5, -4.5, -4.5],
       [-1.5, -1.5, -1.5],
       [ 1.5,  1.5,  1.5],
       [ 4.5,  4.5,  4.5]])
```

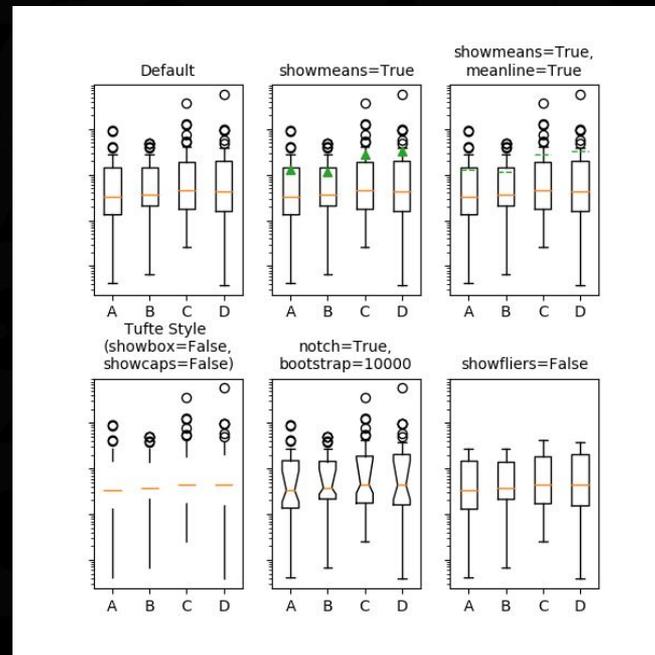
MATPLOTLIB E SEABORN

Biblioteca importante para criar visualizações e gráficos em 2D. A biblioteca já conta nativamente com gráficos de linha, barra, dispersão, boxplot, beeswarm, histogramas entre outros para plotar de maneira simples. O Seaborn utiliza o matplotlib mas tem foco em trabalhar na estética e visualizações mais voltadas para a análise exploratória de dados. Paralelamente o que se utiliza em R: ggplot

link para documentação:

<https://matplotlib.org/2.0.2/index.html>

<https://seaborn.pydata.org/#>



SCIKIT LEARN

Pacote focado em compilar funções e classes relativas a machine learning. Esse pacote é essencial, já que a implementação de algoritmos de regressão, classificação, clusterização entre outros demandam uma complexidade na elaboração. Paralelamente o que se utiliza em R: caret

link para documentação:

<https://scikit-learn.org/stable/>



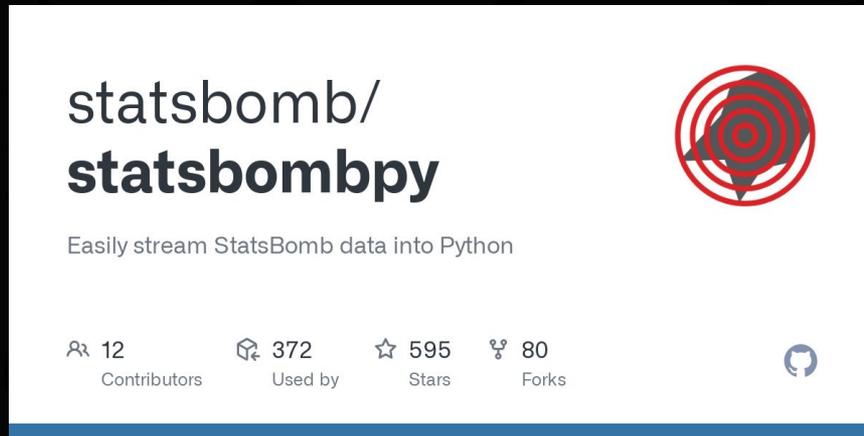
**BIBLIOTECA
S FUTEBOL**

statsbombpy

Biblioteca para interagir com a API de dados do Statsbomb. Ela simplifica o acesso a conjunto de dados, permite verificar dados de competição, temporadas, partidas e jogadores. Simplifica o trabalho de limpeza e construção de pipelines para lidar com dados do Statsbomb.

link para documentação:

<https://github.com/statsbomb/statsbombpy>



statsbomb/
statsbombpy

Easily stream StatsBomb data into Python

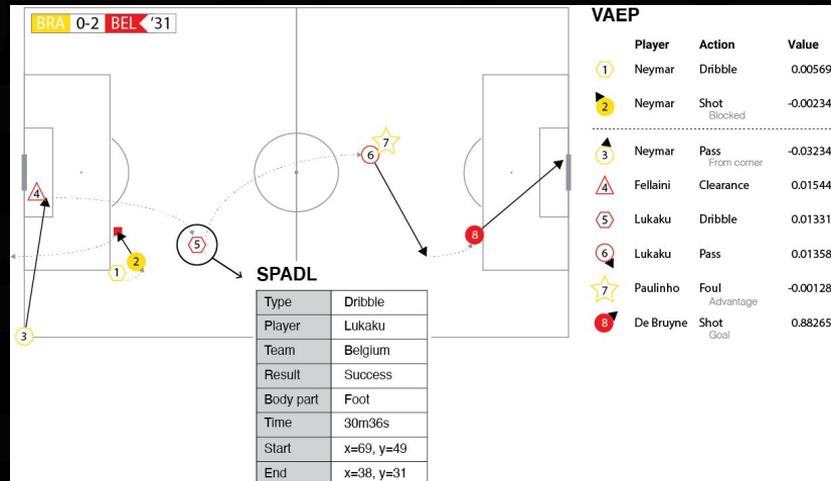
12 Contributors 372 Used by 595 Stars 80 Forks

socceraction

Biblioteca focada em lidar com tasks para modelagem e análise de dados no futebol. Desenvolvida pelos pesquisadores da KULeuven, ela prepara dados numa estrutura de SPADL que é importante para implementação do framework VAEP, na qual identifica o valor e probabilidades associadas às ações executadas por jogadores. Ela é compatível com diferentes provedores de dados.

link para documentação:

<https://socceraction.readthedocs.io/en/latest/index.html>



kloppy

Biblioteca focada em processar e padronizar dados de evento e de tracking de diferentes provedores de dados. Simplifica o trabalho de limpeza, lida com possíveis erros de processamento e fornece um framework para sincronização de dados de evento com tracking.

link para documentação:

<https://kloppy.pysport.org/>



mplsoccer

Biblioteca baseada em matplotlib que é destinada a criação automática da figura de um campo de futebol para visualização de dados específicas. Inclui na documentação visualizações como mapa de calor, redes de passes, mapa de ações de jogadores, sonar de passe, radar e gráficos de pizza.

link para documentação:

<https://mplsoccer.readthedocs.io/en/latest/>

LINGUAGENS DE PROGRAMAÇÃO



Premier League 2019/20 week-wise standings:
A comparison between **Liverpool**, **Manchester City** and **Manchester United**



soccerdata

Biblioteca com foco em limpeza de dados de futebol de diferentes fontes de dados abertos como Elo, soFifa, Understat, FotMoB. Além da raspagem de dados, já possui funções para organizar a leitura dos dados chaves de cada fornecedor.

link para documentação:

<https://soccerdata.readthedocs.io/en/latest/index.htm>

!



SoccerData

Efficiently scrape soccer data from various sources

Dúvidas?

REFERÊNCIAS UTILIZADAS

<https://soccerdata.readthedocs.io/en/latest/index.html>

<https://mplsoccer.readthedocs.io/en/latest/>

<https://kloppy.pysport.org/>

<https://github.com/statsbomb/statsbombpy>

<https://scikit-learn.org/stable/>

<https://pandas.pydata.org/docs/index.html>

<https://socceraction.readthedocs.io/en/latest/index.html>

OBRIGADO!



@CcBatatinha



[linkedin.com/in/caio-batatinha/](https://www.linkedin.com/in/caio-batatinha/)



caio.batatinha@footure.com.br
caiocbatatinha@gmail.com