



WBA0753_v1.0

Linguagens de programação para ciência de dados (*Python com Spark*)



Manipulação de dados com *Python*

Apresentação das bibliotecas. Importação de dados. Acesso e manipulação dos dados.

Bloco 1

Yuri Sá



► Manipulação de dados em *Python*

Vamos nos aproveitar da extensibilidade do *Python* e utilizar as bibliotecas que são o padrão da indústria em manipulação de dados.

Bibliotecas a serem instaladas:

- Numpy.
- Pandas.
- Scipy.

Figura 1 - Logo *Python*



Fonte:

https://upload.wikimedia.org/wikipedia/commons/f/f8/Python_logo_and_wordmark.svg.

Acesso em: 22 mar. 2021.



➤ Manipulação de dados em *Python - Numpy*

A biblioteca *Numpy* fornece algumas estruturas de funcionalidades importantes, que não são padrão da linguagem, tais como:

- *Arrays* e matrizes multidimensionais.
- Operações matemáticas para execução em *arrays* e matrizes.
- Funções de manipulação e conversão de *arrays* e matrizes.

(pip install numpy)

Figura 2 - Logo *NumPy*



Fonte:

https://commons.wikimedia.org/wiki/File:NumPy_logo_2020.svg. Acesso em: 22 mar. 2021.



➤ Manipulação de dados em *Python* - *SciPy*

A biblioteca *SciPy* se apoia em diversas outras bibliotecas (exemplo: *NumPy*) para fornecer um conjunto de funcionalidades para matemática, engenharia e computação científica em geral.

Contém módulos de otimização, álgebra linear, visualização de dados etc.

(pip install scipy)

Figura 3 - Logo *SciPy*



Fonte: <https://www.fullstackpython.com/scipy-numpy.html>. Acesso em: 22 mar. 2021.

➤ Manipulação de dados em *Python - Pandas*

A biblioteca *Pandas* é muito utilizada em análise de dados, fornecendo estruturas e ferramentas para organização e manipulação de dados.

Implementa o *data frame*, uma estrutura de dados com uma organização sequencial e em formato de painel (*Panel Data* – Origem do nome *pandas*).

Também fornece ferramentas para ler, interpretar e escrever dados tabulares.

(pip install pandas)

Figura 4 – Logo *Pandas*



Fonte:

[https://pt.wikipedia.org/wiki/Pandas_\(software\)#/media/Ficheiro:Pandas_logo.svg](https://pt.wikipedia.org/wiki/Pandas_(software)#/media/Ficheiro:Pandas_logo.svg). Acesso em: 22 mar. 2021.

Manipulação de dados com *Python*

Apresentação das bibliotecas. Importação de dados. Acesso e manipulação dos dados .

Bloco 2

Yuri Sá



➤ Manipulação de dados em *Python* – Carregando dados

Carregaremos dados numéricos e tabulares em *python* e demonstraremos, utilizando os pacotes referenciados.

Guia:

- Carregar um CSV.
- Exibir as colunas como *arrays*.
- Fazer operações básicas com os *arrays*.
- Plotar em gráficos.
- Criar e manipular *data frames*.



Teoria em Prática

Bloco 3

Yuri Sá



➤ Reflita sobre a seguinte situação

Uma empresa precisa importar os dados sobre clientes que estão armazenados em um arquivo CSV (*Comma Separated Values*), com várias colunas de forma ordenada.

Que pacote e função devem ser utilizados?



► Norte para a resolução...

- Avalie qual o melhor pacote que oferece estruturas de dados para tal importação.
- Procure, na documentação do pacote, qual a função para importar dados ordenadamente.



Dica do (a) Professor (a)

Bloco 4

Yuri Sá



Ache o editor ideal para você

Utilize IDEs ou editores de código nos quais se sinta confortável e ágil. Sua produtividade dependerá deste editor.

Procure softwares que tenham funcionalidades que sejam mais úteis a você, seja execução interativa, área de plotagem, comentários avançados, notebooks etc.

O universo é vasto e é fácil encontrar soluções diversas.





Referências

DOCS SCIPY. SciPy. **Docs SciPy**, dezoito de fevereiro de dois mil e vinte um. Disponível em:
<https://docs.scipy.org/doc/scipy/reference/>. Acesso em: 22 mar. 2021.

NUMPY. NumPy v1.20 Manual. **NumPy**, trinta e um de janeiro de dois mil e vinte um. Disponível em:
<https://numpy.org/doc/stable/>. Acesso em: 22 mar. 2021.

PANDAS PYDATA. Pandas documentation. **Pandas**, dois de março de dois mil e vinte e um. Disponível em:
<https://pandas.pydata.org/docs/>. Acesso em: 22 mar. 2021.



Bons estudos!

