

Senac

Todas as formas
de aprender

tecno
logia

CURSO

**_BIG
DATA
SCIENCE**

_ANALISTA DE DADOS



ALESSANDRO VIEIRA

RECAPITULANDO...

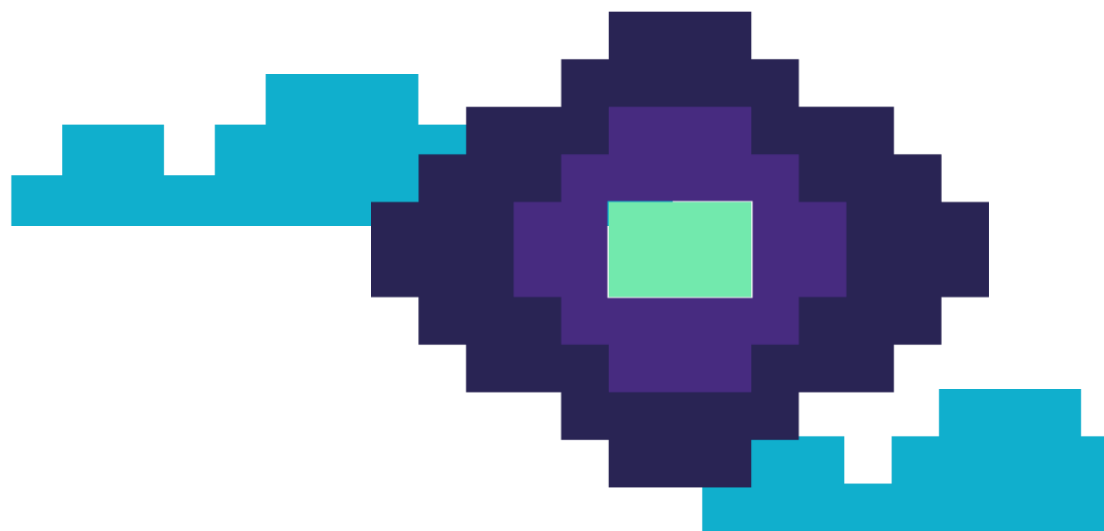
Antes de mais nada, vamos recapitular!

**Quais são as suas dúvidas
sobre a aula anterior?**

AULA

02

Introdução à Análise de Dados



ANALISTA DE DADOS – BIG DATA SCIENCE

A **Ciência de Dados** é um ramo que vem ganhando cada vez mais notoriedade, várias empresas de pequeno a grande porte, como a Netflix, Airbnb e Google já possuem atividades de tomada de decisão baseadas em dados.

Nesse cenário, a linguagem **Python** é bastante utilizada devido a sua versatilidade e simplicidade, contando com uma vasta quantidade de bibliotecas, e entre elas, o **Pandas**, uma das ferramentas essenciais quando se fala em Ciência de Dados.

Nesta apresentação, vamos conhecer a biblioteca **Pandas**, entender sobre as suas estruturas básicas, e também como instalar a ferramenta.

ANALISTA DE DADOS – BIG DATA SCIENCE

O que é Pandas?

É uma biblioteca para Ciência de Dados de código aberto (open source), construída sobre a linguagem **Python**, e que providencia uma abordagem rápida e flexível, com estruturas robustas para se trabalhar com dados relacionais (ou rotulados), e tudo isso de maneira simples e intuitiva.

De maneira geral, o **Pandas** pode ser utilizado para várias atividades e processos, entre eles: limpeza e tratamento de dados, análise exploratória de dados (EDA), suporte em atividades de Machine Learning, consultas e queries em bancos de dados relacionais, visualização de dados e muito mais.

E além disso, também possui ótima integração com várias outras bibliotecas muito utilizadas em Ciência de Dados, tais como: **Numpy**, **Scikit-Learn**, **Seaborn**, **Altair**, **Matplotlib**, **Plotly** e outros.

ANALISTA DE DADOS – BIG DATA SCIENCE

Instalando o Pandas

Uma maneira comum de instalar o Pandas é utilizando o PIP, o sistema de gerenciamento de pacotes do Python.

pip install pandas

Particularmente eu prefiro criar um ambiente virtual antes da instalação da biblioteca, pois dessa forma consigo trabalhar com versões diferentes da biblioteca.

python -m venv <nome_do_ambiente>

Para ativarmos o ambiente virtual devemos executar as seguintes linhas de comando:

cd <nome_do_ambiente>

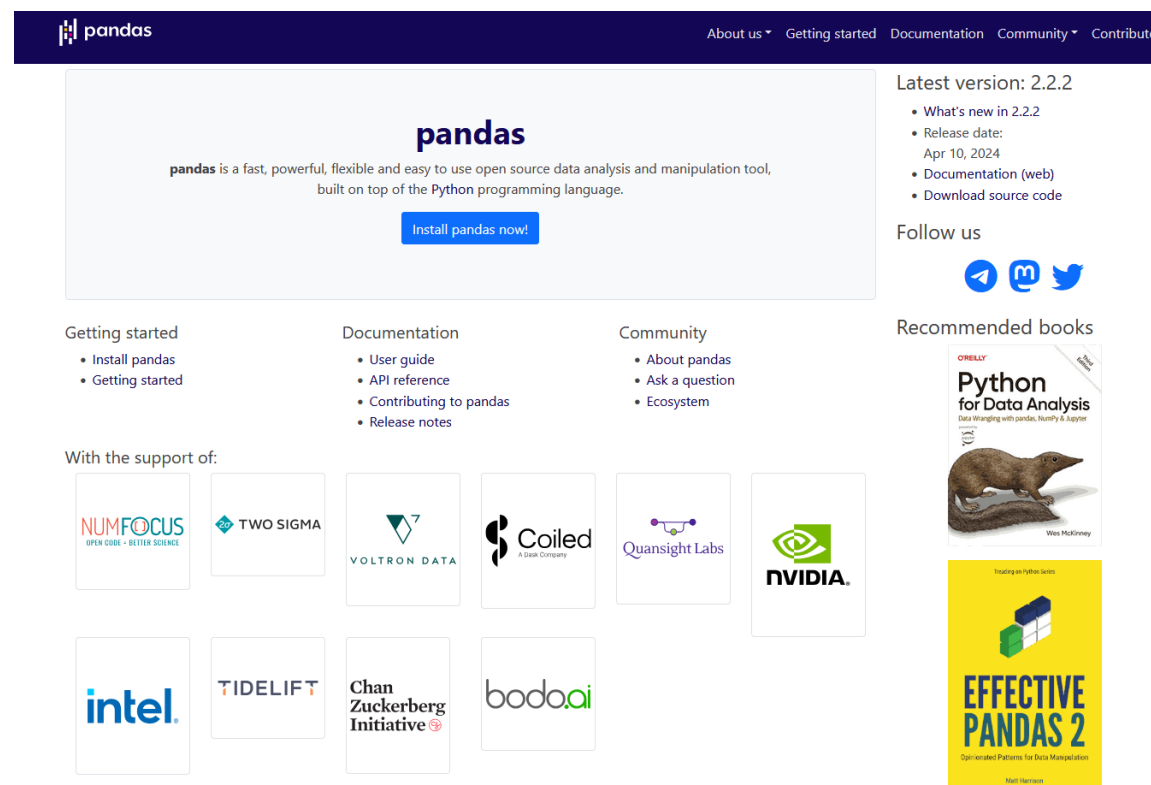
cd scripts

activate

ANALISTA DE DADOS – BIG DATA SCIENCE

Para visualizar a documentação dessa biblioteca, devemos acessar o seguinte link:

<https://pandas.pydata.org/>



The screenshot shows the pandas website homepage. At the top is a dark blue navigation bar with the pandas logo and links for 'About us', 'Getting started', 'Documentation', 'Community', and 'Contribute'. The main content area has a light gray background. On the left, the word 'pandas' is prominently displayed, followed by a description: 'pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.' Below this is a blue button that says 'Install pandas now!'. To the right of this section, the 'Latest version: 2.2.2' is noted, along with a list of links: 'What's new in 2.2.2', 'Release date: Apr 10, 2024', 'Documentation (web)', and 'Download source code'. Below the main content, there are three columns of links: 'Getting started' (Install pandas, Getting started), 'Documentation' (User guide, API reference, Contributing to pandas, Release notes), and 'Community' (About pandas, Ask a question, Ecosystem). Further down, a section titled 'With the support of:' displays a grid of logos from various sponsors, including NUMFOCUS, TWO SIGMA, VOLTRON DATA, Coiled, Quansight Labs, NVIDIA, intel, TIDELIFT, Chan Zuckerberg Initiative, and bodo.ai. On the far right, under the heading 'Recommended books', two book covers are shown: 'Python for Data Analysis' by Wes McKinney and 'EFFECTIVE PANDAS 2' by Mark Harris.

ANALISTA DE DADOS – BIG DATA SCIENCE

Estrutura de Dados - Listas

Uma **Lista** (ou array) é uma estrutura de dados em Python que pode conter múltiplos elementos de diferentes tipos, como números, strings, ou até mesmo outras listas. As listas são mutáveis, ou seja, podem ser modificadas após a sua criação.

```
lista = [23, 'São Paulo', True, [1, 2, 3]]
```


ANALISTA DE DADOS – BIG DATA SCIENCE

Estrutura de Dados - Séries

Uma **Série** é uma estrutura de dados unidimensional, semelhante a uma coluna em uma tabela. Cada elemento da Série tem um índice associado, que pode ser numérico ou não. No Pandas, uma Série é um array unidimensional com rótulos.

```
idades = pd.Series([23, 35, 45], index=['Ana', 'Bruno', 'Carlos'])
```

ANALISTA DE DADOS – BIG DATA SCIENCE

Estrutura de Dados - DataFrames

Um **DataFrame** é uma estrutura de dados bidimensional, semelhante a uma tabela em uma base de dados ou uma planilha do Excel. Ele é composto por linhas e colunas, onde cada coluna pode conter diferentes tipos de dados (numéricos, strings, etc.).

No contexto do Pandas, um **DataFrame** é uma coleção de Séries, onde cada Série representa uma coluna.

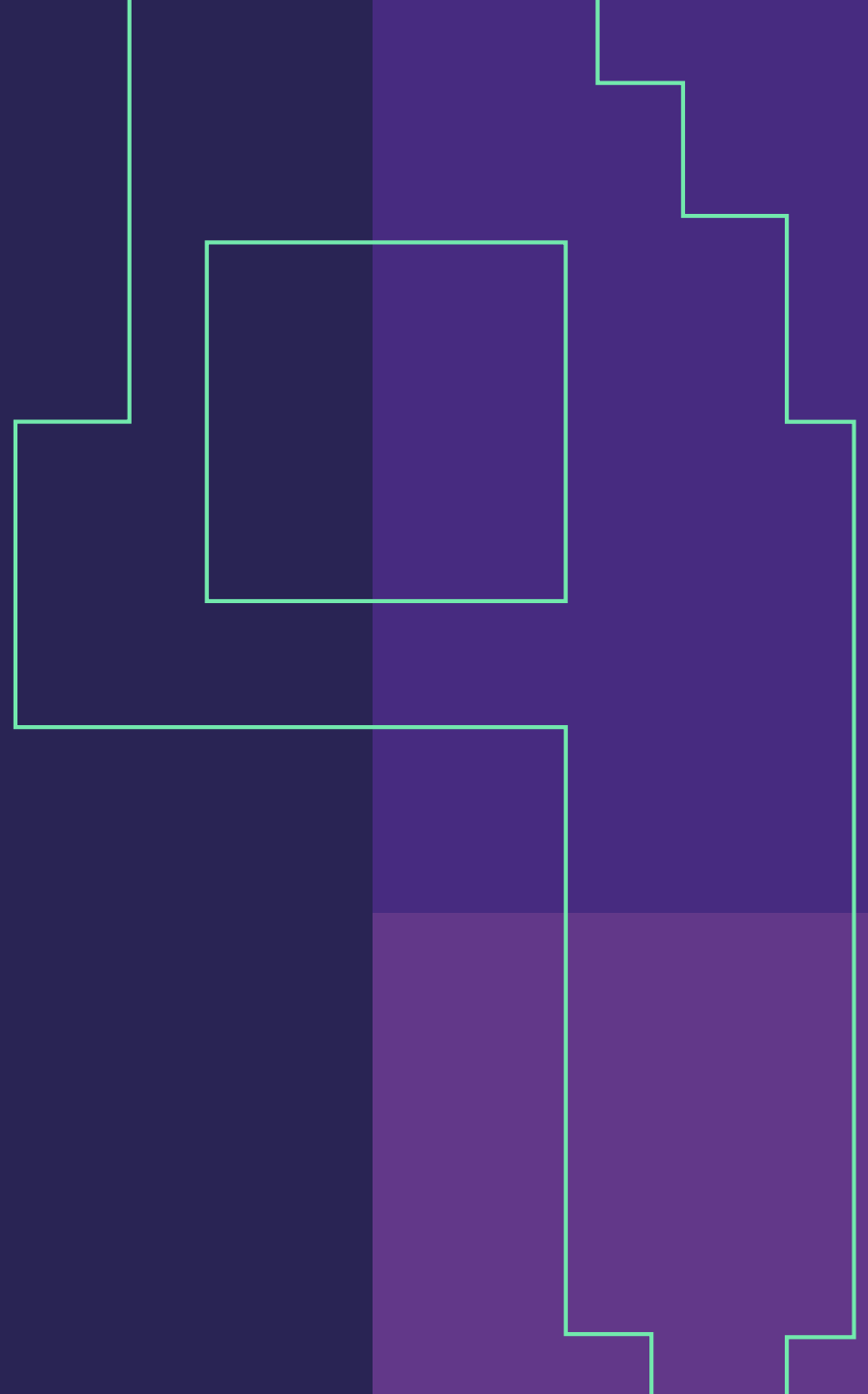
```
data = {  
    'Nome': ['Ana', 'Bruno', 'Carlos'],  
    'Idade': [23, 35, 45],  
    'Cidade': ['São Paulo', 'Rio de Janeiro', 'Belo Horizonte']  
}  
  
df = pd.DataFrame(data)
```



BORA **PRATICAR?**



Senac
tecnologia



ANALISTA DE DADOS – BIG DATA SCIENCE

Praticando

1- Faça um programa que leia duas séries com 10 números inteiros cada e ao final mostre a soma, a subtração, a multiplicação e a divisão entre elas.

2- Escreva um programa que leia uma série com 10 idades e ao final separe as que estão iguais e acima de 18 anos das que estão abaixo de 18 anos.

ANALISTA DE DADOS – BIG DATA SCIENCE

Praticando

3- O Delegado responsável pela Delegacia de roubos e furtos de automóveis, entrou em contato com você e te solicitou um auxílio, para obter duas informações:

- A quantidade de roubos de automóveis + furto de automóveis diária, dos últimos 7 dias.
- A taxa de recuperação de automóveis diária, dos últimos 7 dias, sabendo que para se chegar a esse número, deve-se dividir a quantidade de recuperação de automóveis pela quantidade de roubo de automóveis.

Ele te enviou os seguintes dados:

- Roubo de automóveis: 100,90,80,120,110,90,70
- Furto de automóveis: 80,60,70,60,100,50,30
- Recuperação de automóveis: 70,50,90,80,100,70,50

E pediu para apresentar na tela o resultado das informações solicitadas.

**Chegamos ao
final do nosso
encontro.**

Até o próximo.



Não esqueça de reservar um momento do seu dia
para fazer a **JORNADA DE FIXAÇÃO**.