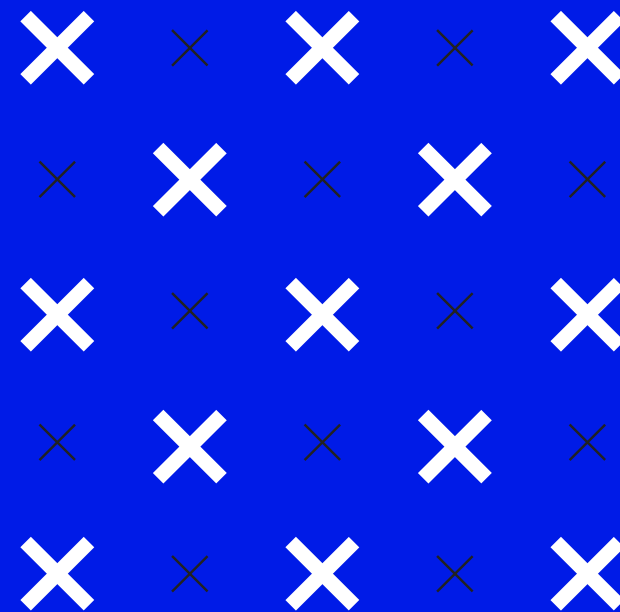


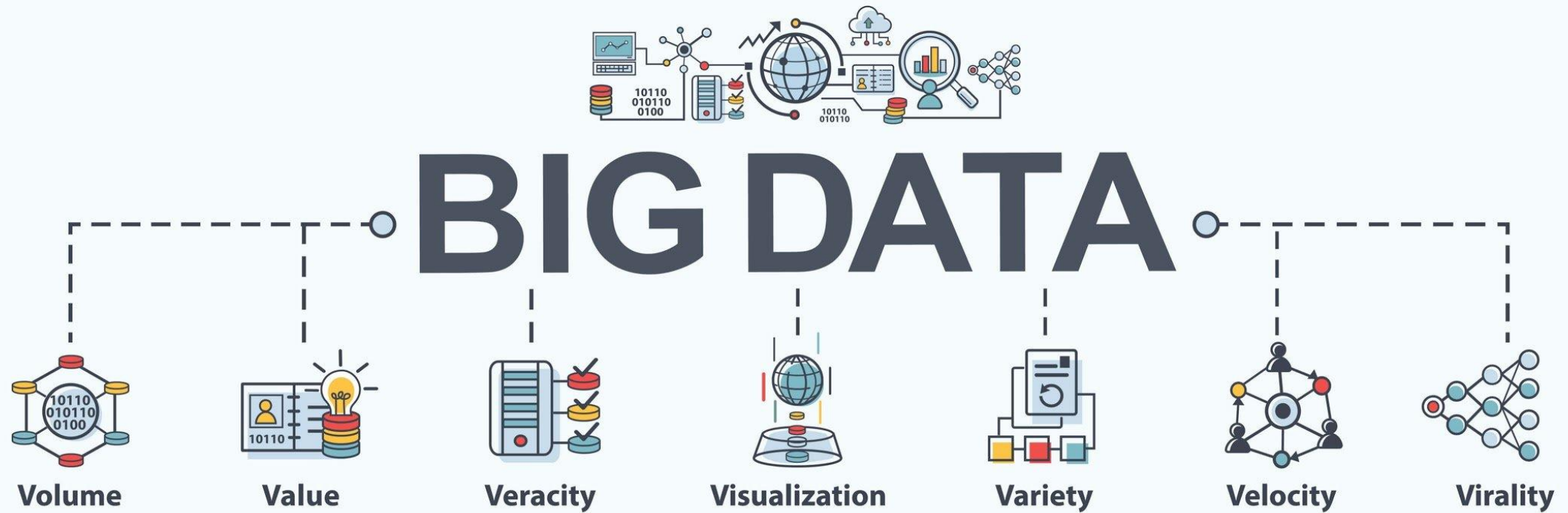


ANÁLISE E VISUALIZAÇÃO DE DADOS

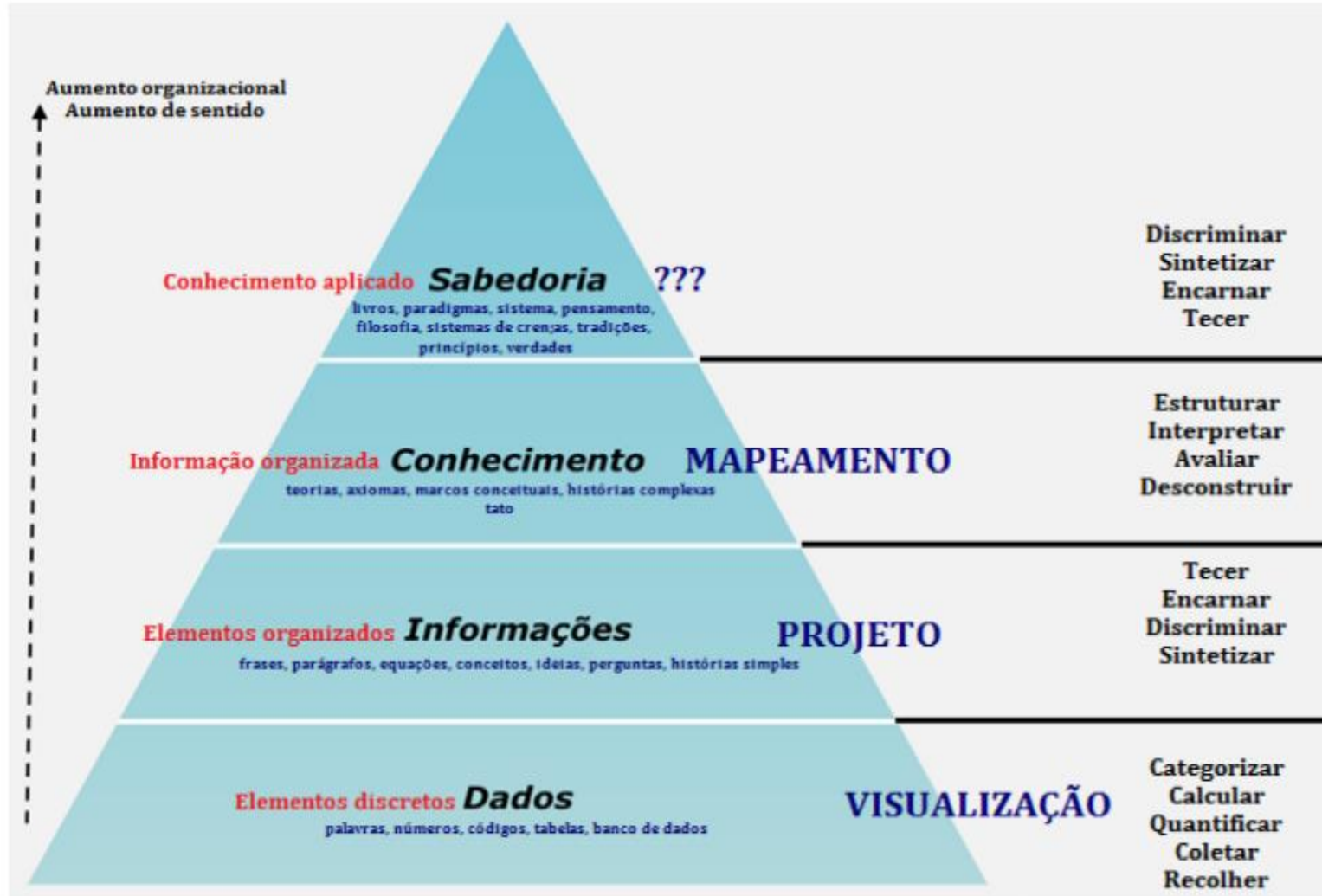


mentorama.
@prof.felipeassuncao

Volume de dados



Dado, informação, conhecimento e sabedoria



Dado, informação, conhecimento e sabedoria



Cientista de Dados

- O **cientista de dados** trabalha com a coleta, o armazenamento e o tratamento de dados, transformando tudo isso em inteligência de negócio
- Faz o levantamento e organização de um volume grande de informações, criando relatórios e tornando esses dados inteligíveis para as pessoas na tomada de decisões

Análise e Visualização de Dados

Análise e visualização de dados



Neste módulo

Aula 1 - Numpy

Aula 2 - Pandas

Aula 3 - Matplotlib

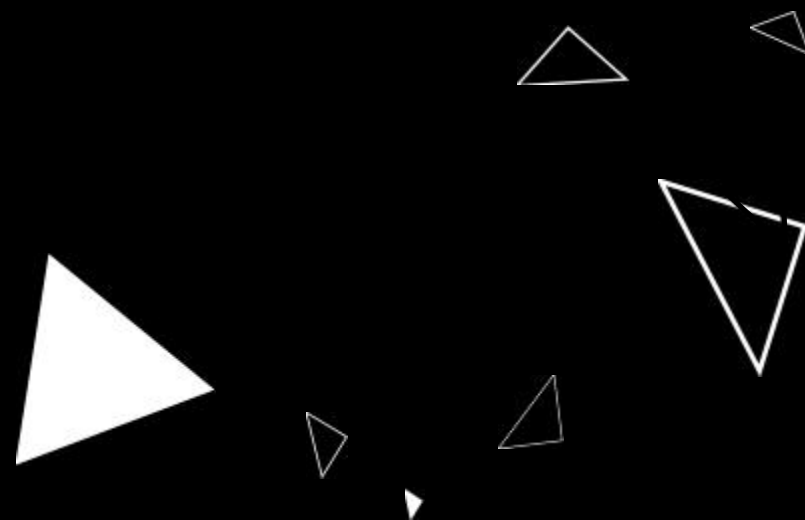
Aula 4 - Projeto

Recursos e ferramentas

- Editor de código de sua preferência

1. NUMPY

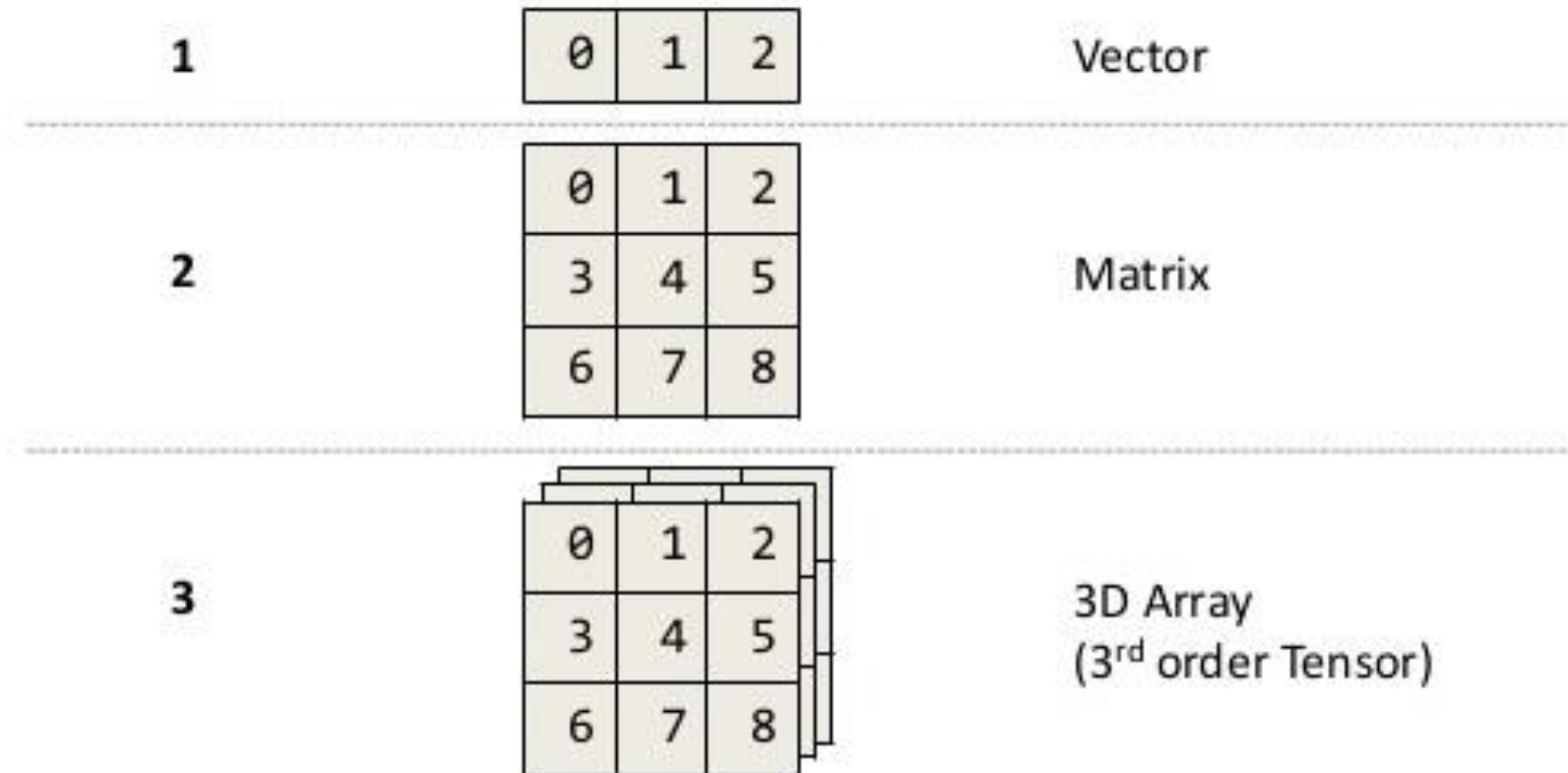
mentorama.



Numpy

- NumPy é um pacote para a linguagem Python que suporta arrays e matrizes multidimensionais, possuindo uma larga coleção de funções matemáticas para trabalhar com estas estruturas.
- Os arrays são as principais estruturas de dados do numpy
- Arrays são objetos semelhantes a listas mas muito mais rápidos de se processar
- Através dos arrays podemos criar vetores (1D) e matrizes(2D) e tensor / matrizes (3D)

Numpy arrays



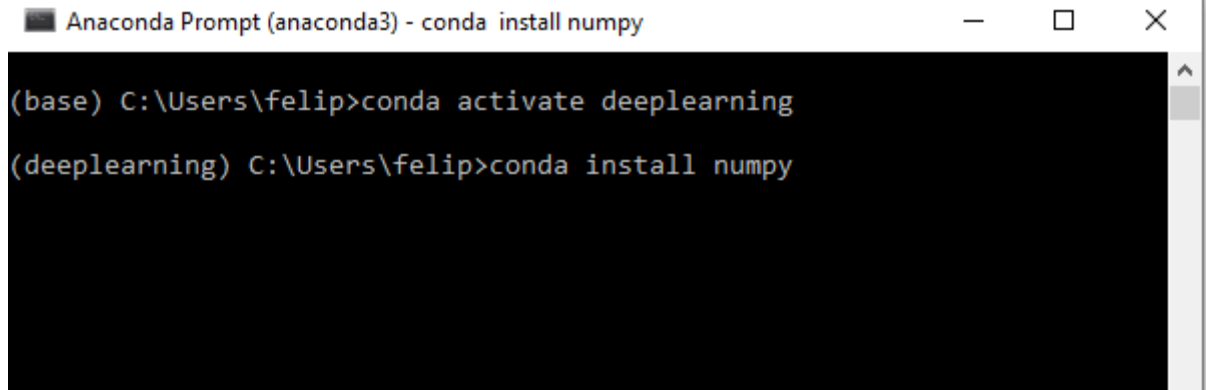
NumPy

- Se você está trabalhando com o Anaconda, use o seguinte código no seu terminal:

```
conda install numpy  
pip install numpy
```

```
>>> import numpy as np
```

mentorama.



```
Anaconda Prompt (anaconda3) - conda install numpy  
  
(base) C:\Users\felip>conda activate deeplearning  
(deeplearning) C:\Users\felip>conda install numpy
```

Para instalar os pacotes do Anaconda, acesse:
<https://anaconda.org/anaconda/repo>



Vamos praticar?

- Nesta prática iremos explorar a utilização do módulo NumPy e os seus principais comandos



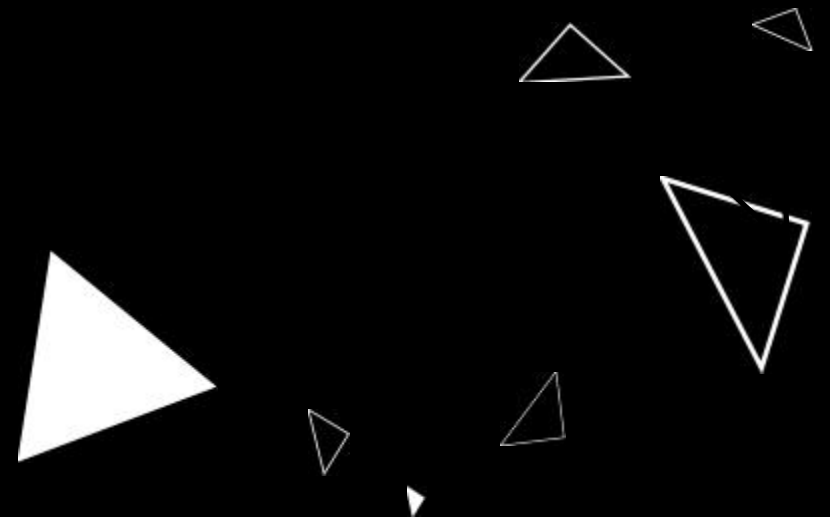
Resumo

- O que é o NumPy
- Principais estruturas de dados do NumPy
- Exemplos práticos



2. PANDAS

mentorama.



Pandas

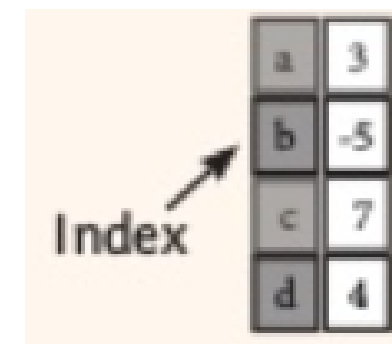
- É uma biblioteca de computação científica para Python que juntamente com Numpy, fornece estrutura de dados e ferramentas para análise de dados em Python

```
>>> import pandas as pd
```



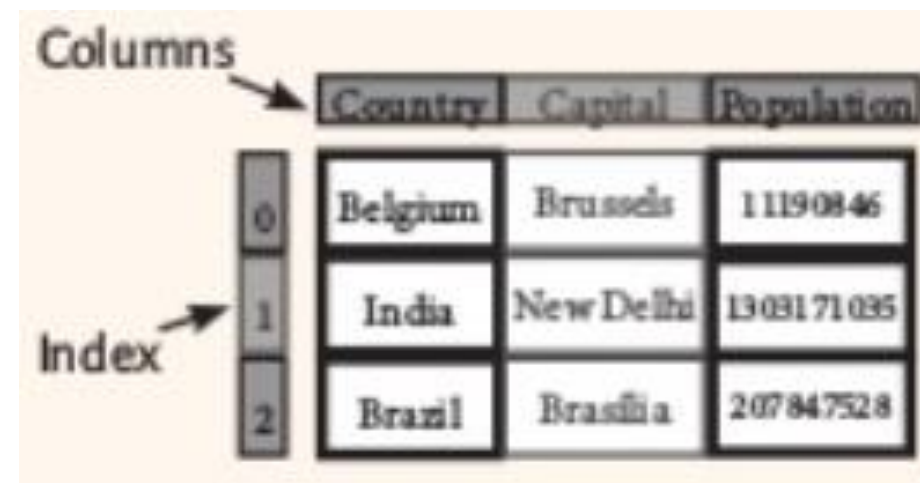
Estrutura de dados no Pandas

- **Séries:** Um array unidimensional rotulado capaz de fixar qualquer tipo de dados
- **Data Frame:** Uma estrutura de dados de duas dimensões com colunas de tipos potencialmente diferentes



A diagram illustrating a Pandas Series. It consists of a vertical column of four cells. The first cell contains 'a' and the second contains 'b'. An arrow labeled 'Index' points to the 'b' cell. The third cell contains 'c' and the fourth contains 'd'. To the right of these cells are numerical values: 3, -5, 7, and 4, each aligned with its corresponding letter.

a	3
b	-5
c	7
d	4



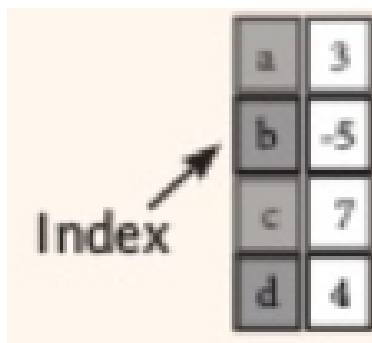
A diagram illustrating a Pandas Data Frame. It is a 2D table with three columns and three rows. The columns are labeled 'Country', 'Capital', and 'Population'. The rows are indexed 0, 1, and 2. An arrow labeled 'Columns' points to the header row, and an arrow labeled 'Index' points to the first column of data.

	Country	Capital	Population
0	Belgium	Brussels	11190846
1	India	New Delhi	1303171035
2	Brazil	Brasilia	207847528

Estrutura de dados no Pandas

Declarando uma série:

```
>>> s = pd.Series ([3, -5, 7, 4], index =  
['a','b','c','d'])
```



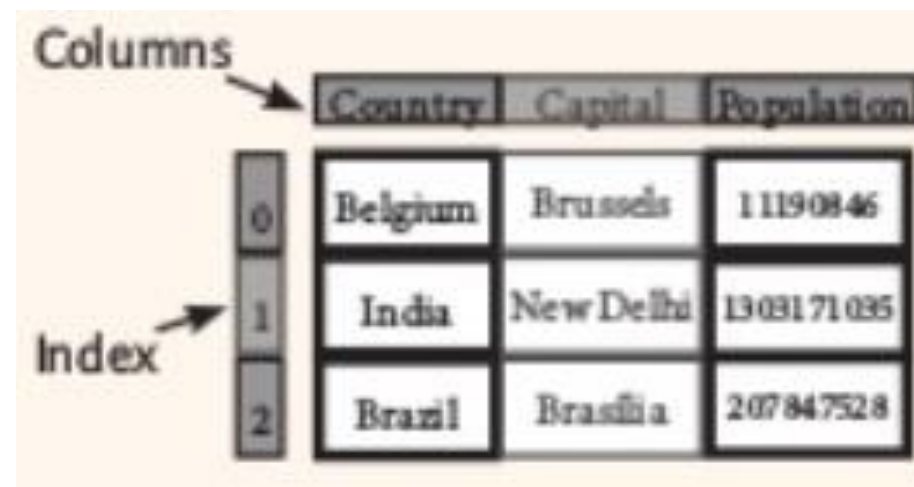
A diagram illustrating a Pandas Series. It consists of a vertical column of four cells. The first cell contains 'a' and the second contains 'b', both in grey boxes. The third cell contains 'c' and the fourth contains 'd', both in white boxes. To the right of these cells are the values 3, -5, 7, and 4, respectively. An arrow labeled 'Index' points to the first cell containing 'a'.

a	3
b	-5
c	7
d	4

Declarando um data frame:

```
>>> Data = ['Country': ['Belgium', 'India', 'Brazil'],  
            'Capital':['Brussels', 'New Delhi', 'Brasilia'],  
            'Population': [111190846, 1303171035,  
                           207847528])
```

```
>>> df = pd.DataFrame(data, columns  
=['Country', 'Capital', 'Population'])
```

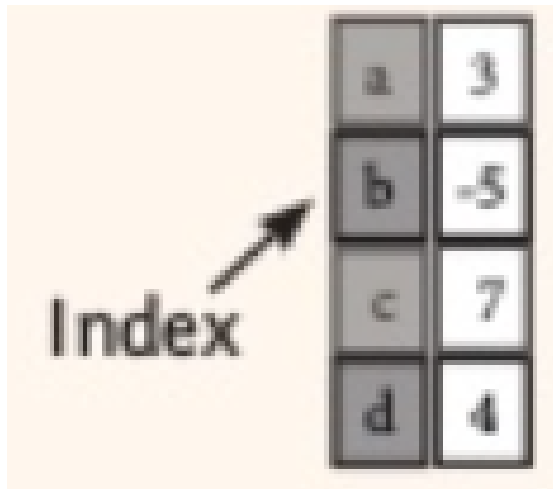


A diagram illustrating a Pandas DataFrame. It shows a table with three columns: 'Country', 'Capital', and 'Population'. The first column has three rows: 'Belgium', 'India', and 'Brazil'. The second column has three rows: 'Brussels', 'New Delhi', and 'Brasilia'. The third column has three rows: '11190846', '1303171035', and '207847528'. An arrow labeled 'Columns' points to the top row of the table. An arrow labeled 'Index' points to the first column of the table.

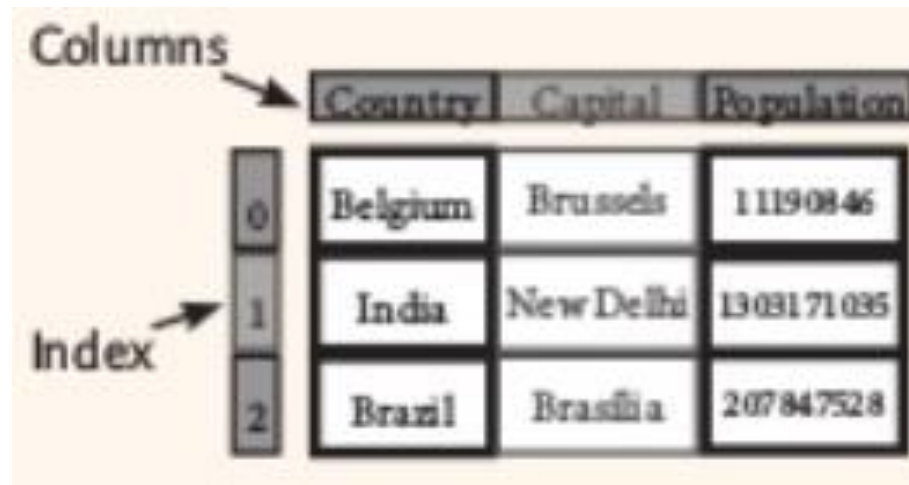
	Country	Capital	Population
0	Belgium	Brussels	11190846
1	India	New Delhi	1303171035
2	Brazil	Brasilia	207847528

Trabalhando com dados

- Ler e escrever para CSV
- Ler e escrever para xlsx (Excel)



a	3
b	-5
c	7
d	4



Columns	Country	Capital	Population	
Index	0	Belgium	Brussels	11190846
1	India	New Delhi	1303171035	
2	Brazil	Brasilia	207847528	

Vamos praticar?

- Nesta prática iremos explorar a utilização do Pandas e os seus principais comandos



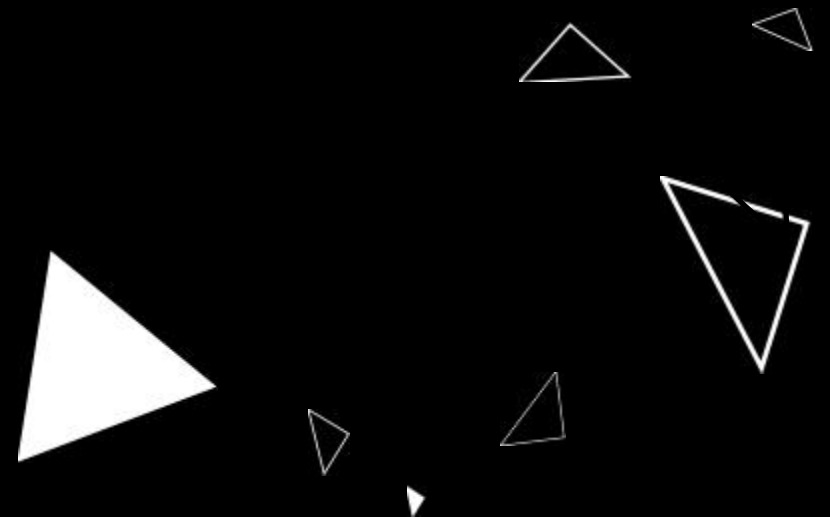
Resumo

- Trabalhando com Séries
- Trabalhando com DataFrames
- Trabalhando com Base de Dados
- Exemplos práticos



3.MATPLOTLIB

mentorama.



Matplotlib

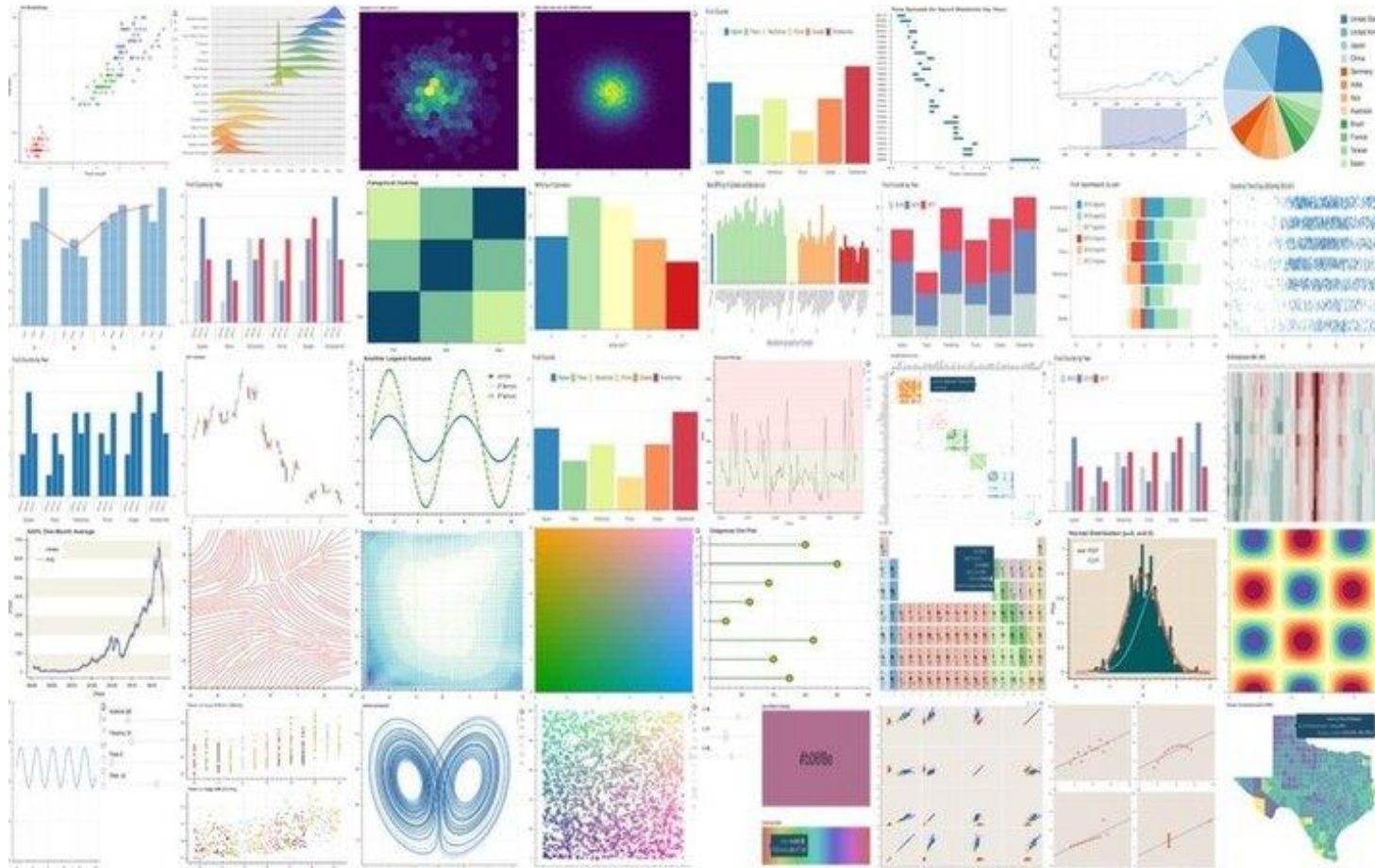
- Matplotlib é uma biblioteca de software para criação de gráficos e visualizações de dados em geral, feita para o Python e para extensão de matemática NumPy

Visualização de dados com o pyplot

- O pyplot é um módulo do Matplotlib para criação de gráficos.
- Para utilizá-lo é necessário fazer a importação:

```
>>> import matplotlib.pyplot as plt
```


Análise e visualização de dados



Vamos praticar?

- Nesta prática iremos explorar a utilização do Matplotlib e seus principais comandos



Resumo

- Configurações iniciais dos gráficos
- Visualização de dados
- Exemplos de gráficos



PROJETO

mentorama.

