

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**

Bibliotecas: Pandas, NumPy, SciPy, Matplotlib e Seaborn

Pandas – Ler e escrever

Aula 1

Código da aula: [DADOS]ANO1C2B3S24A1

**Bibliotecas: Pandas,
NumPy, SciPy,
Matplotlib e Seaborn**

Mapa da Unidade 5 Componente 3

NumPy – Manipulação
de array

semana

22

Pandas – Acesso e
seleção

semana

23

semana

20

NumPy: acesso,
inspeção e índice

semana

21

NumPy – Operação
de arrays

semana

24

Você está aqui!
Pandas – Ler e
escrever

**Bibliotecas: Pandas,
NumPy, SciPy,
Matplotlib e Seaborn**

Mapa da Unidade 5 Componente 3

Você está aqui!

Pandas – Ler e escrever

Aula 1

Código da aula: [DADOS]ANO1C2B3S24A1

24



Objetivos da aula

- Salvar arquivos a partir da biblioteca Pandas do Python.



Recursos didáticos

- Recurso audiovisual para exibição de vídeos e imagens.
- Acesso ao laboratório de informática e/ou à internet.
- Software Anaconda/Jupyter Notebook instalado ou similar.



Duração da aula

50 minutos



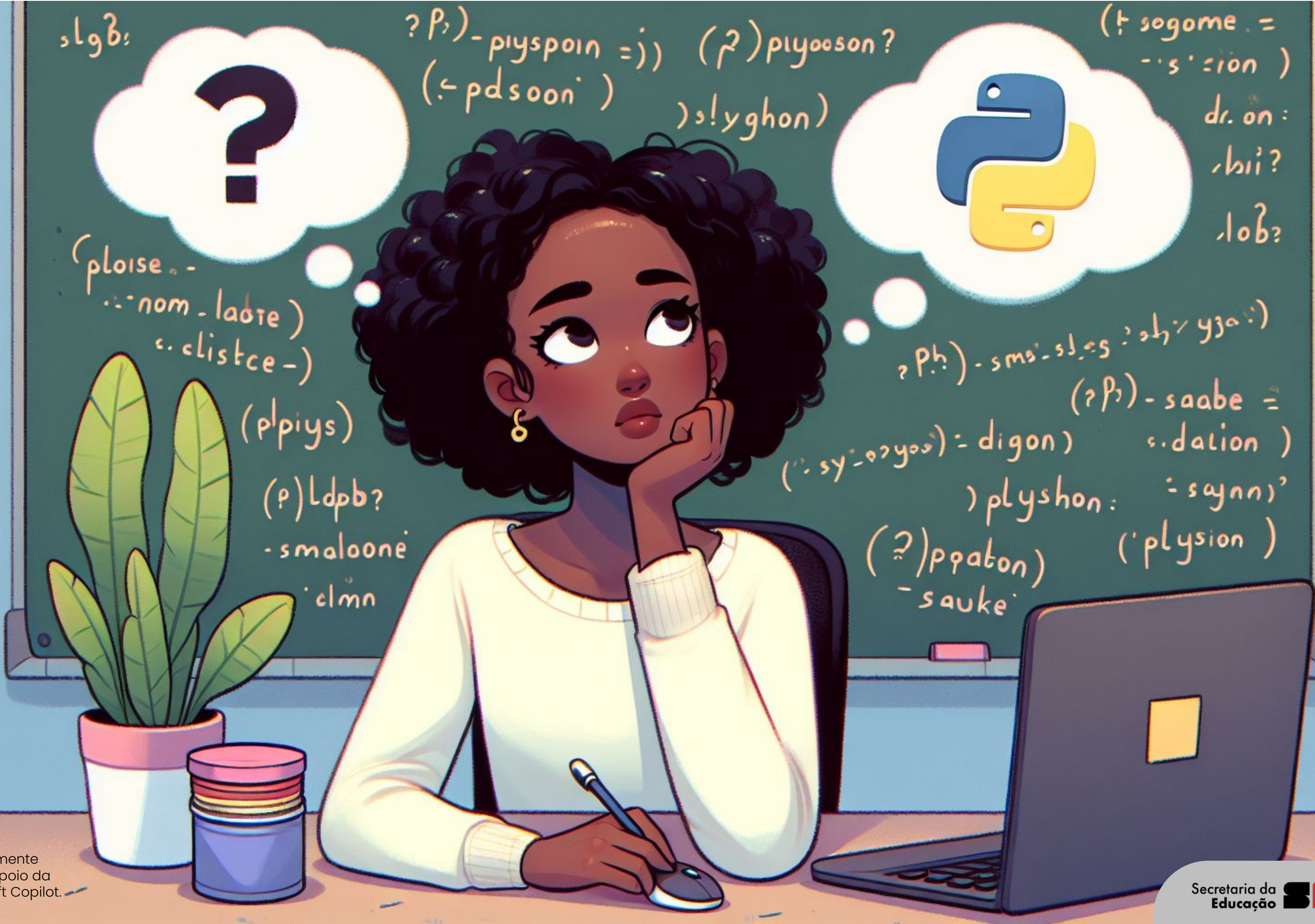
Competências técnicas

- Ser proficiente em linguagens de programação para manipular e analisar grandes conjuntos de dados.
- Usar técnicas para explorar e analisar dados, aplicar modelos estatísticos, identificar padrões, realizar inferências e tomar decisões baseadas em evidências.



Competências socioemocionais

- Colaborar efetivamente com outros profissionais, como cientistas de dados e engenheiros de dados.
- Trabalhar em equipes multifuncionais, colaborando com colegas, gestores e clientes.



Primeiras
ideias

O que a imagem anterior expressa?

- Já vimos que o Pandas cria dataframes, mas onde salvar este arquivo?
- Será que o Pandas do Python também lê um arquivo?
- Será que eu consigo ler ou salvar um arquivo do tipo xlsx (Excel)?

Ponto de partida

Salvar um arquivo no Pandas

Temos um dataframe do Pandas na nossa IDE. Como salvar o dataframe a seguir para poder usar novamente mais tarde?

	Nome	C1 - Introdução a Ciência de Dados, Ferramentas e Jargões da Área	C2 - Ética e Responsabilidade em Inteligência Artificial	C3 - Programação Aplicada a Ciência de Dados	C4 - Lógica, Algoritmo e Operação de Planilhas Eletrônicas	C5 - Banco de Dados e Computação em Nuvem
0	Micael Miguel Ehenvic Puzuho de Oliveira	6.0	3.0	7.0	4.0	6.0
1	Janaína Mupru Froiri	9.0	2.0	6.0	7.0	4.0
2	Álvaro Savo da Silva	3.0	7.0	7.0	2.0	5.0
3	Marlon Rosatto Neto	4.0	1.0	7.0	5.0	1.0
4	Sandra Inetirn	4.0	0.0	9.0	5.0	8.0
5	Charlene Sandra Klein de Simões	0.0	9.0	2.0	6.0	3.0
6	Gilberto Martinho Riga	8.0	2.0	4.0	2.0	6.0
7	Gilson de Albuquerque	4.0	8.0	6.0	1.0	3.0
8	Fred Jair dos Santos	8.0	1.0	9.0	8.0	9.0
9	Damião Gowa de Soares	4.0	1.0	3.0	6.0	7.0
10	Filipe Anakin de Padilha	2.0	0.0	3.0	1.0	7.0
11	Guiomar Mourão	3.0	1.0	5.0	5.0	9.0

Elaborado especialmente para o curso.

Pandas

O Pandas é uma **biblioteca de código aberto** para a linguagem de programação Python, criada por Wes McKinney em 2008. Ele foi projetado especificamente para lidar com estruturas de dados e análise de dados.

- ▶ Oferece alta performance e ferramentas de análise de dados.
- ▶ É amplamente utilizado por cientistas e analistas de dados.
- ▶ É essencial para limpeza, transformação, agregação e visualização de dados em Python.

Pandas

As principais estruturas de dados são **DataFrames** e **Series**.

DataFrames

- São como tabelas com colunas representadas por Series.
- Bidimensional, como tabelas com colunas representadas por Series.

Series

- São estruturas unidimensionais associadas a rótulos (índices).
- Unidimensional, armazena qualquer tipo de dado.

Construindo o **conceito**

Pandas

Criando o dataframe a seguir:

```
1 import pandas as pd
2
3 data = {"Nome": ["Ana", "Jorge", "Carol", "Mateus"],
4         "Idade": [25, 30, 22, 15],
5         "Cidade": ["Santos", "São Paulo", "Campinas", "Bauru"]}
6 meu_dataframe = pd.DataFrame(data)
7 meu_dataframe
```

	Nome	Idade	Cidade
0	Ana	25	Santos
1	Jorge	30	São Paulo
2	Carol	22	Campinas
3	Mateus	15	Bauru

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Pandas – Salvar

Um arquivo pode conter diversas extensões. As extensões de arquivos com dados tabulares são importantes para identificar o tipo de arquivo e garantir que ele seja aberto no software apropriado.

Aqui estão dois formatos comuns:

► CSV (Comma-Separated Values)

Extensão: .csv

Descrição: Armazena dados tabulares em texto simples, com valores separados por vírgulas.

Uso: Compatível com muitos programas e linguagens de programação.

► XLSX (Excel Workbook)

Extensão: .xlsx

Descrição: Formato baseado em XML padrão para o Excel 2010 e versões posteriores.

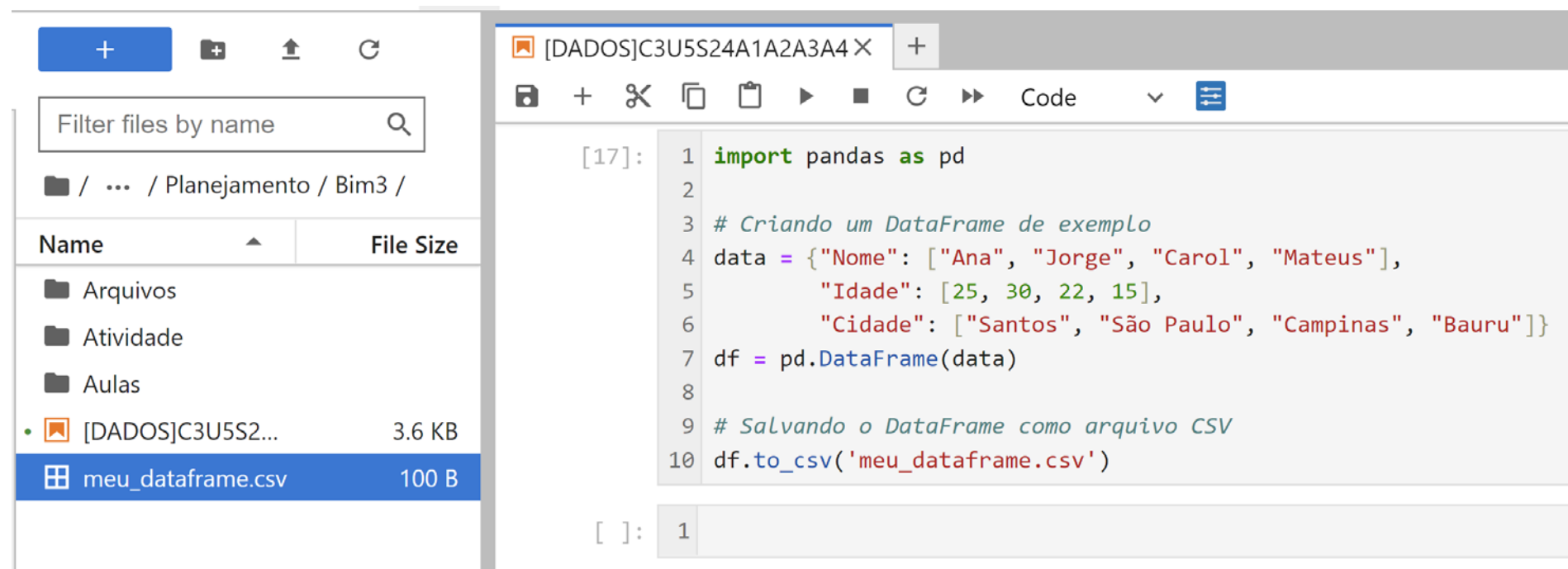
Recursos: Suporta várias planilhas, formatação, fórmulas e gráficos.

Construindo o conceito

Pandas – Salvar

Salvar como arquivo CSV: o método **to_csv()** permite salvar um DataFrame como um arquivo CSV.

- Salvar no diretório de trabalho.



```
[17]: 1 import pandas as pd
      2
      3 # Criando um DataFrame de exemplo
      4 data = {"Nome": ["Ana", "Jorge", "Carol", "Mateus"],
      5        "Idade": [25, 30, 22, 15],
      6        "Cidade": ["Santos", "São Paulo", "Campinas", "Bauru"]}
      7 df = pd.DataFrame(data)
      8
      9 # Salvando o DataFrame como arquivo CSV
     10 df.to_csv('meu_dataframe.csv')
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



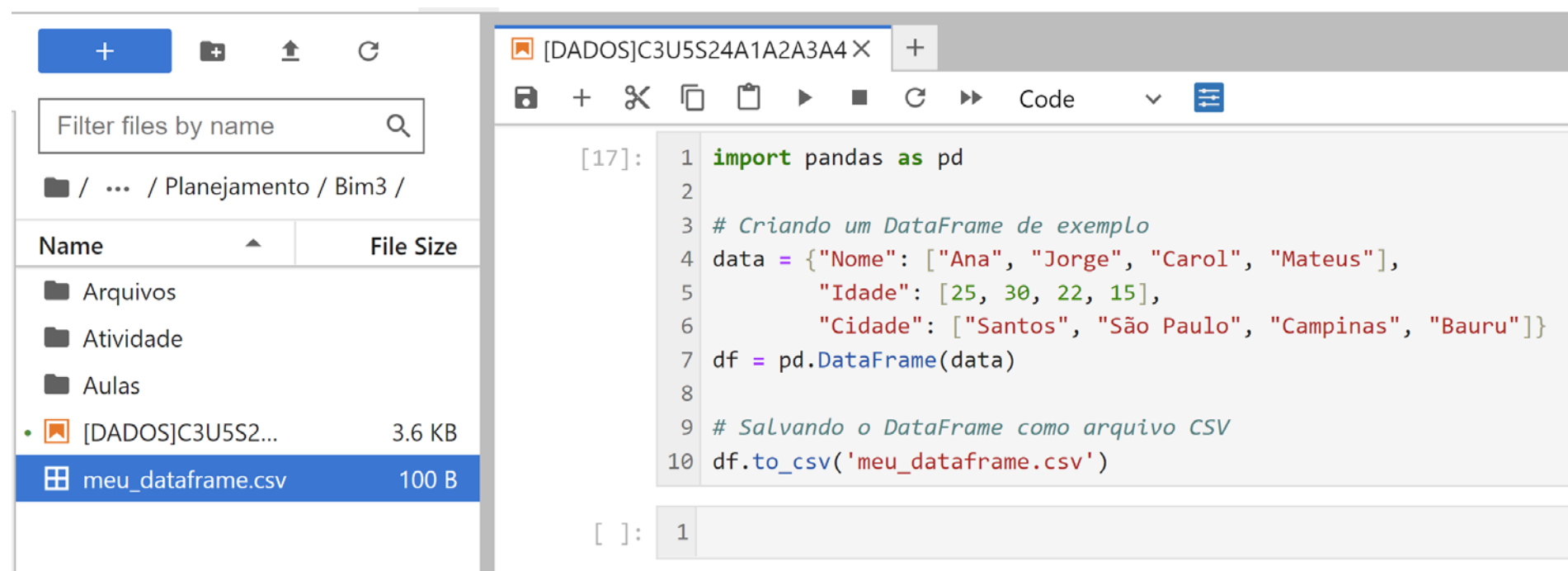
Tome nota:

Observe que foi criado um arquivo com extensão csv na mesma pasta em que você criou o arquivo Jupyter Notebook.

Construindo o conceito

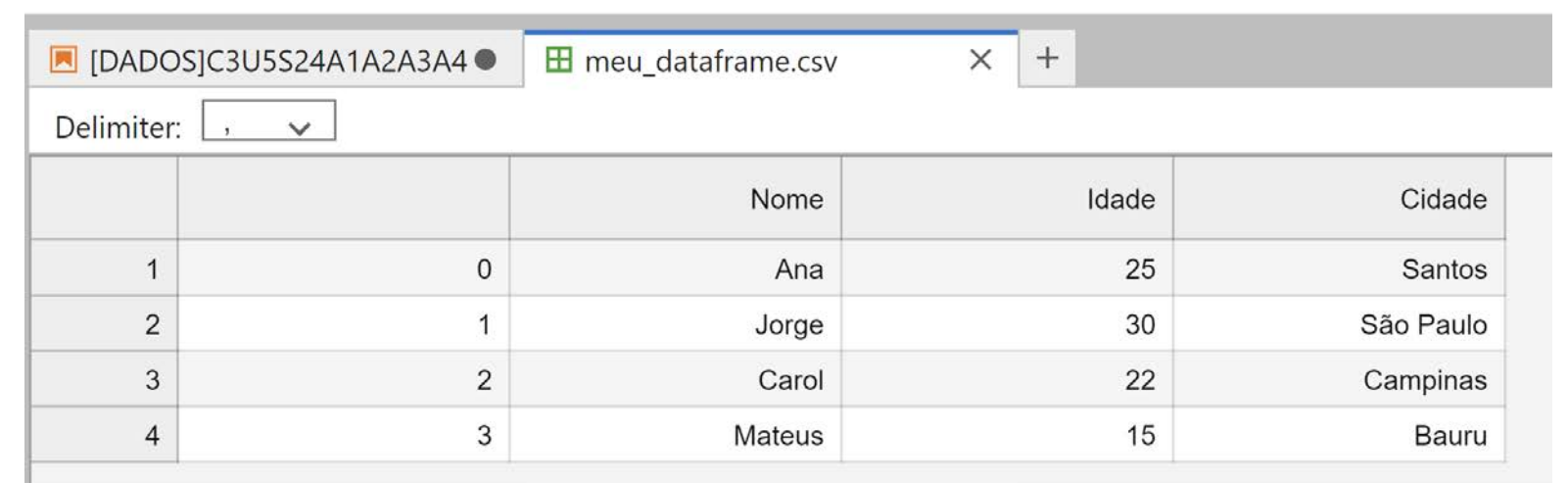
Pandas – Salvar

Como fica o arquivo salvo?



The screenshot shows the Jupyter Notebook interface. On the left, the file explorer displays a directory structure with folders 'Arquivos', 'Atividade', and 'Aulas'. Below them, a file named '[DADOS]C3U5S24A1A2A3A4' is listed with a size of 3.6 KB. The main area shows the code execution output for cell [17]:

```
[17]: 1 import pandas as pd
      2
      3 # Criando um DataFrame de exemplo
      4 data = {"Nome": ["Ana", "Jorge", "Carol", "Mateus"],
      5        "Idade": [25, 30, 22, 15],
      6        "Cidade": ["Santos", "São Paulo", "Campinas", "Bauru"]}
      7 df = pd.DataFrame(data)
      8
      9 # Salvando o DataFrame como arquivo CSV
     10 df.to_csv('meu_dataframe.csv')
```



The screenshot shows the Jupyter Notebook interface with the file explorer on the left displaying the saved file 'meu_dataframe.csv' with a size of 100 B. The main area shows the table view of the CSV file, with the delimiter set to comma (,).

		Nome	Idade	Cidade
1	0	Ana	25	Santos
2	1	Jorge	30	São Paulo
3	2	Carol	22	Campinas
4	3	Mateus	15	Bauru

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



Tome nota:

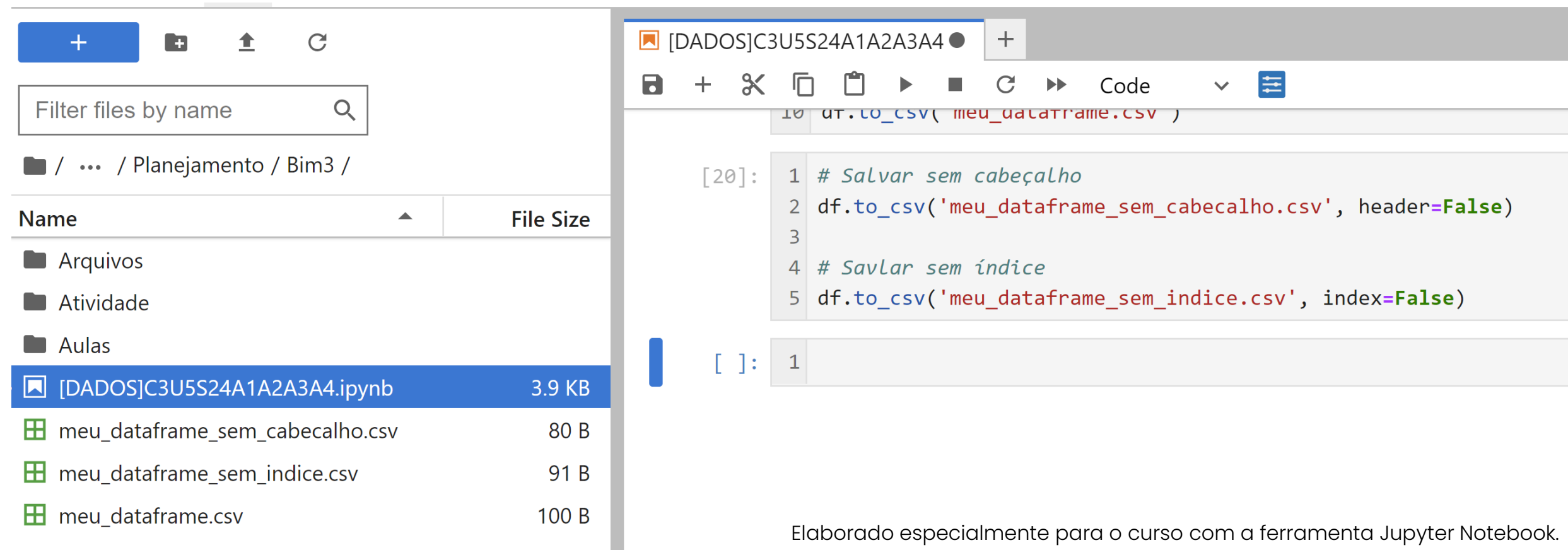
Observe que foi criada uma coluna antes da coluna Nome. Essa coluna é uma coluna de índice.

Construindo o conceito

Pandas – Salvar

Salvar como arquivo CSV:

- Salvar sem cabeçalho e índice.



The screenshot displays a Jupyter Notebook interface. On the left, the file explorer shows a directory structure with folders 'Arquivos', 'Atividade', and 'Aulas'. Below them, a list of files is shown with their names and sizes:

Name	File Size
[DADOS]C3U5S24A1A2A3A4.ipynb	3.9 KB
meu_dataframe_sem_cabecalho.csv	80 B
meu_dataframe_sem_indice.csv	91 B
meu_dataframe.csv	100 B

On the right, the code editor shows the following Python code:

```
[20]: 1 # Salvar sem cabeçalho
      2 df.to_csv('meu_dataframe_sem_cabecalho.csv', header=False)
      3
      4 # Salvar sem índice
      5 df.to_csv('meu_dataframe_sem_indice.csv', index=False)
```

Below the code, the output of the execution is shown as a single line: 1.

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



Atenção:

Se o arquivo já existir, o método `to_csv` irá substituir o arquivo existente.



Construindo
o **conceito**

Pandas – Salvar

Como fica o arquivo salvo?

The screenshot shows a Jupyter Notebook interface. On the left, a file browser sidebar displays a directory structure: 'Planejamento / Bim3 /'. Below this, a table lists files and folders:


Name	File Size
Arquivos	
Atividade	
Aulas	
[DADOS]C3U5S24A1A2A3A4.ipynb	3.9 KB
meu_dataframe_sem_cabecalho.csv	80 B
meu_dataframe_sem_indice.csv	91 B
meu_dataframe.csv	100 B


The main area shows a Jupyter Notebook cell with the following code:


```
[20]: 1 # Salvar sem cabeçalho
      2 df.to_csv('meu_dataframe_sem_cabecalho.csv', header=False)
      3
      4 # Salvar sem índice
      5 df.to_csv('meu_dataframe_sem_indice.csv', index=False)
```

The output of the cell is shown below the code:

```
[ ]: 1
```

 [DADOS]C3U5S24A1A2A3A4 X

 meu_dataframe_sem_indice.c X

 meu_data

Delimiter: , ▼

	Nome	Idade	Cidade
1	Ana	25	Santos
2	Jorge	30	São Paulo
3	Carol	22	Campinas
4	Mateus	15	Bauru

[DADOS]C3U5S24A1A2A3A4 X

meu_dataframe_sem_indice.c X

meu_dataframe_sem_cabeca X

+

Delimiter:

,

▼

	0	Ana	25	Santos
1	1	Jorge	30	São Paulo
2	2	Carol	22	Campinas
3	3	Mateus	15	Bauru

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Pandas – Salvar

Salvar como arquivo CSV:

- Salvar em um lugar específico.

```
1 # Salvar em um lugar específico
2
3 df.to_csv('C:\Users\Admin\Desktop\meu_dataframe_local.csv', index=False)
```

Caminho (*path*):
pasta na qual quero
salvar o arquivo

Nome do arquivo

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Pandas – Salvar

Salvar como arquivo Excel:

O método **to_excel()** salva um DataFrame como um arquivo Excel.

```
1 # Salvar um DataFrame como arquivo Excel
2
3 df.to_excel('meu_dataframe.xlsx', index=False)
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook

```
%pip install openpyxl
```

```
Collecting openpyxl
  Downloading openpyxl-3.1.2-py2.py3-none-any.whl.metadata (2.5 kB)
Collecting et_xmlfile (from openpyxl)
  Downloading et_xmlfile-1.1.0-py3-none-any.whl.metadata (1.8 kB)
Downloading openpyxl-3.1.2-py2.py3-none-any.whl (249 kB)
----- 250.0/250.0 kB 1.4 MB/s eta 0:00:00
Downloading et_xmlfile-1.1.0-py3-none-any.whl (4.7 kB)
Installing collected packages: et_xmlfile, openpyxl
Successfully installed et_xmlfile-1.1.0 openpyxl-3.1.2
```



Tome nota:

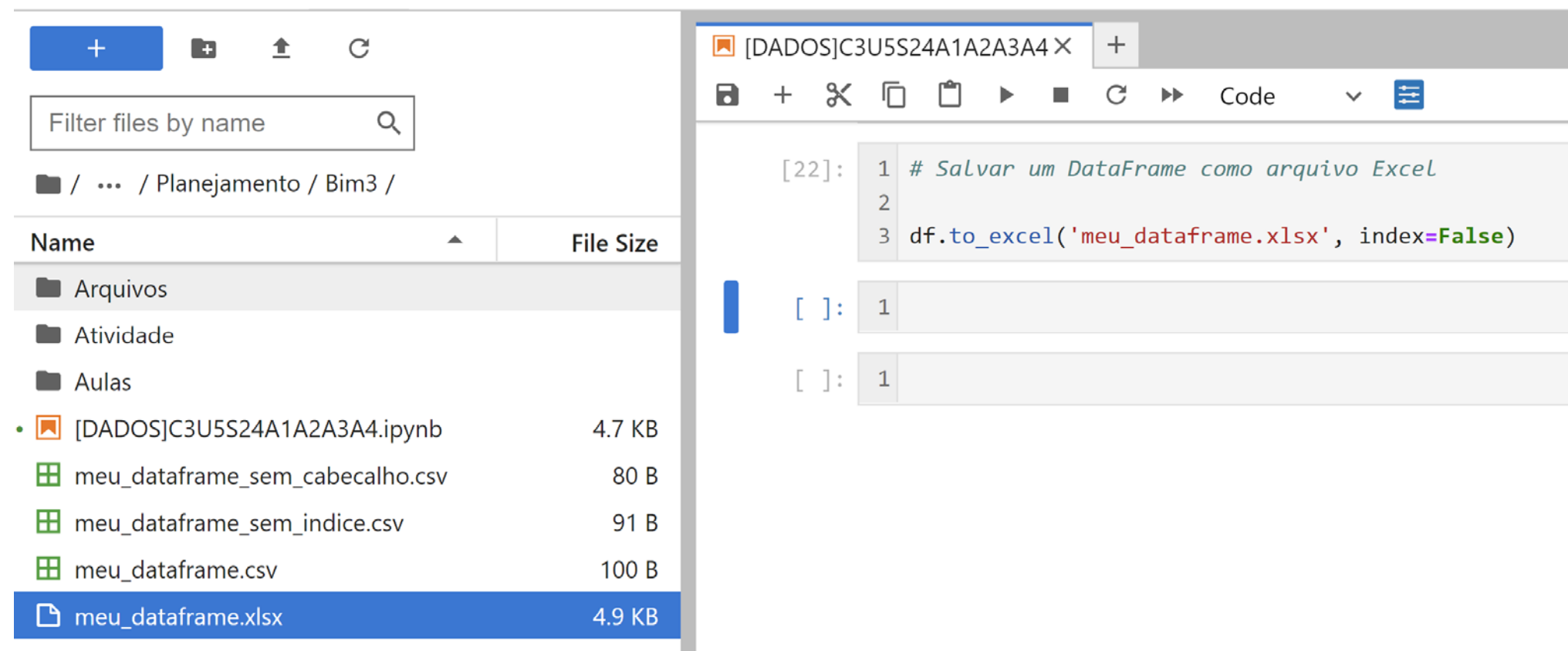
Se o módulo der erro, ou seja, não estiver instalado, vai dar um erro. Para solucionar o problema, digite na célula: %pip install openpyxl

Construindo o conceito

Pandas – Salvar

Salvar como arquivo Excel:

O método **to_excel()** salva um DataFrame como um arquivo Excel.



The screenshot shows a Jupyter Notebook interface. On the left is a file explorer with a search bar 'Filter files by name' and a list of files and folders. The file 'meu_dataframe.xlsx' is selected. The main area shows a code cell with the following code:

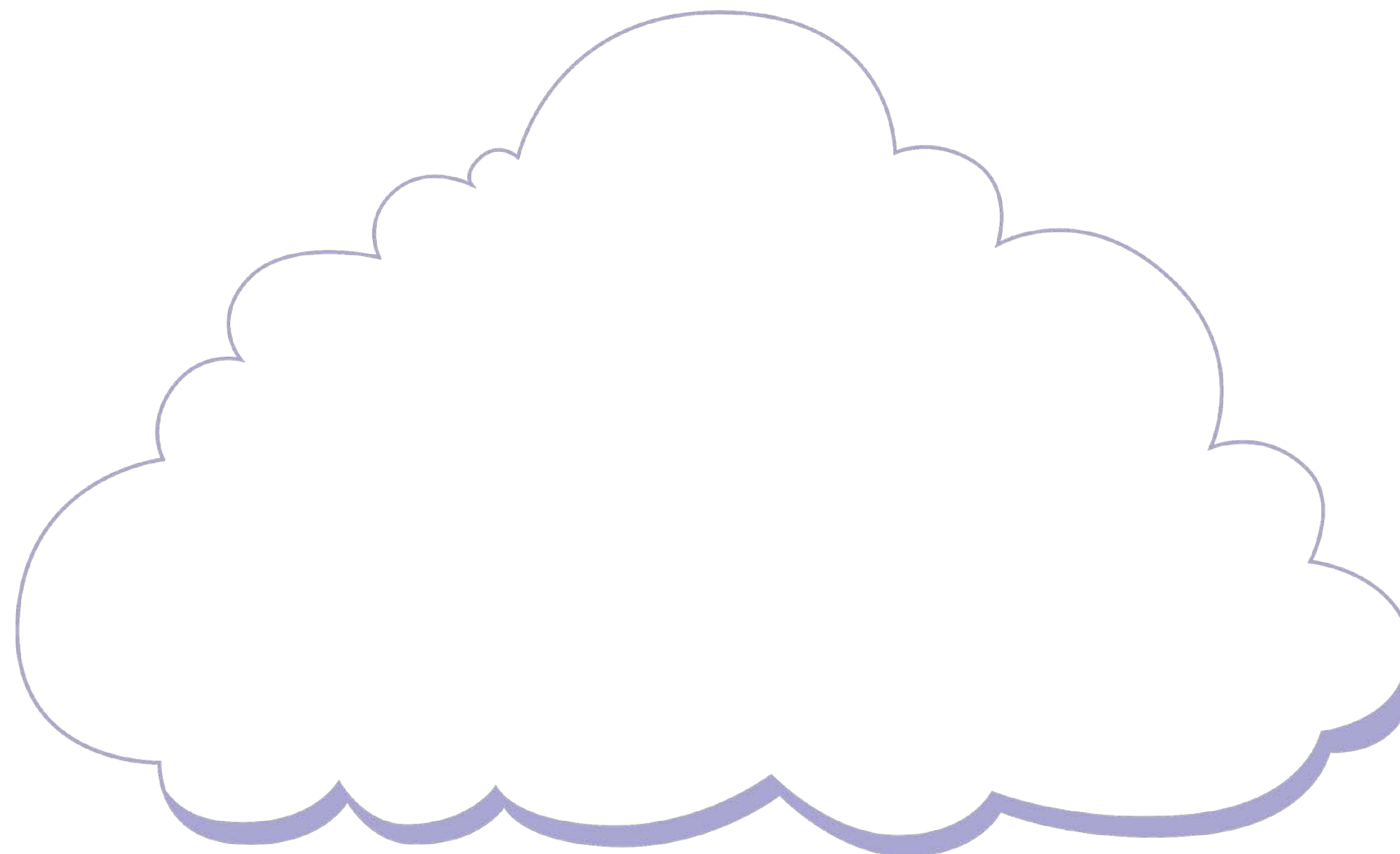
```
[22]: 1 # Salvar um DataFrame como arquivo Excel
      2
      3 df.to_excel('meu_dataframe.xlsx', index=False)
```

Below the code cell, the output is displayed as a single line: 1.

	A	B	C	D
1	Nome	Idade	Cidade	
2	Ana	25	Santos	
3	Jorge	30	São Paulo	
4	Carol	22	Campinas	
5	Mateus	15	Bauru	
6				
-				

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Nuvem de palavras



O que nós
**aprendemos
hoje?**

© Getty Images



© Getty Images

O que nós
aprendemos
hoje?

Então ficamos assim...

- 1** Aprendemos que o Pandas possui o método **to_csv** e **to_excel** para salvar dataframes em arquivos de extensão csv e xlsx. Existem argumentos para personalizar o modo de salvar, por exemplo, sem cabeçalho e/ou sem índice;
- 2** Compreendemos que Dataframes salvos em csv e xlsx podem ser abertos em planilhas eletrônicas, por exemplo, Excel;
- 3** Vimos que é possível salvar um arquivo na pasta local ou escolher qual pasta salvar o arquivo.

Saiba mais

Não quer mais se perder em planilhas gigantescas e dados desorganizados? Aprenda a dominar o Pandas, a biblioteca Python que transforma a análise de dados em um processo rápido e intuitivo!

ALURA. *Pandas: conhecendo a biblioteca*. 01 Apresentação. Disponível em: <https://cursos.alura.com.br/course/pandas-conhecendo-biblioteca/task/125504>. Acesso em: 3 jul. 2024.

Está pronto para dominar o Pandas e turbinar sua análise de dados? O Guia prático do Pandas da Databricks te leva do zero ao expert em DataFrames!

DATABRICKS. *DataFrame do Pandas*, [s.d.]. Disponível em: <https://www.databricks.com/br/glossary/pandas-dataframe#:~:text=Pandas%C3%A9%20uma%20biblioteca%20de,ferramentas%20de%20an%C3%A1lise%20de%20dados>. Acesso em: 3 jul. 2024.

Referências da aula

MCKINNEY, W. *Python para análise de dados: tratamento de dados com Pandas, NumPy e Jupyter*. São Paulo: Novatec, 2023.

PANDAS. *Página inicial*, [s.d.]. Disponível em: <https://pandas.pydata.org/>. Acesso em: 3 jul. 2024.

Identidade visual: Imagens © Getty Images.

Educação Profissional Paulista

Técnico em
**Ciência de
Dados**