

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**

Bibliotecas: Pandas, NumPy, SciPy, Matplotlib e Seaborn

Pandas: acesso e seleção

Aula 2

Código da aula: [DADOS]ANO1C2B3S23A2

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

Mapa da Unidade 5 Componente 3

semana

22

NumPy:
manipulação
de array

semana

24

Pandas: ler e
escrever

semana

23

Você está aqui!
Pandas: acesso
e seleção

semana

25

Pandas: transformando
Data Frame

semana

26

Pandas: combinando
Data Frame

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

Mapa da Unidade 5 Componente 3

Você está aqui!

Pandas: acesso e seleção

Aula 2

Código da aula: [DADOS]ANO1C2B3S23A2

23



Objetivos da aula

- Conhecer o conceito de Data Frame e series do Pandas.



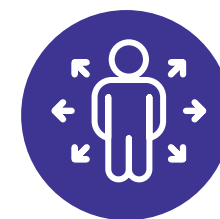
Recursos didáticos

- 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 com base 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.

Construindo
o **conceito**

Pandas

Series:

Uma Series é uma estrutura de dados unidimensional capaz de armazenar qualquer tipo de dados, como inteiros, *strings*, números flutuantes, objetos Python etc.

Cada elemento em uma Series é associado a um rótulo, também conhecido como índice.

Construindo
o **conceito**

Pandas

Data Frame:

Um Data Frame é uma estrutura de dados bidimensional, semelhante a uma tabela de banco de dados ou a uma planilha eletrônica, em que cada coluna é uma Series.

Data Frames oferecem funcionalidades ricas e flexíveis para a manipulação e análise de dados.

Construindo o conceito

Diferença entre Series e Data Frame

A Series apresenta uma estrutura unidimensional, enquanto o Data Frame tem uma estrutura bidimensional.

			Data	Produto	Região	Quantidade	Valor Unitário	Receita
0	2022-12-05							
1	2022-06-26							
2	2022-03-16	0	2022-12-05	Shampoo	Leste	360	104	37440
3	2023-06-13	1	2022-06-26	Shampoo	Sul	578	226	130628
4	2022-06-03							
	...							
3647	2023-07-21	2	2022-03-16	Pasta de dente	Sul	939	119	111741
3648	2023-11-21	3	2023-06-13	Maionese	Norte	472	83	39176
3649	2023-11-22							
3650	2023-06-06	4	2022-06-03	Pasta de dente	Norte	859	31	26629
3651	2023-02-13							

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Semelhança entre Series e Data Frame

Tanto em uma Series quanto em um Data Frame, os índices começam em zero por padrão.

			Data	Produto	Região	Quantidade	Valor Unitário	Receita
0	2022-12-05							
1	2022-06-26							
2	2022-03-16	0	2022-12-05	Shampoo	Leste	360	104	37440
3	2023-06-13	1	2022-06-26	Shampoo	Sul	578	226	130628
4	2022-06-03							
	...							
3647	2023-07-21	2	2022-03-16	Pasta de dente	Sul	939	119	111741
3648	2023-11-21	3	2023-06-13	Maionese	Norte	472	83	39176
3649	2023-11-22							
3650	2023-06-06	4	2022-06-03	Pasta de dente	Norte	859	31	26629
3651	2023-02-13							

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Pandas – Data Frame

Observe a tabela e responda:

	Data	Produto	Região	Quantidade	Valor Unitário	Receita
0	2022-12-05	Shampoo	Leste	360	104	37440
1	2022-06-26	Shampoo	Sul	578	226	130628
2	2022-03-16	Pasta de dente	Sul	939	119	111741
3	2023-06-13	Maionese	Norte	472	83	39176
4	2022-06-03	Pasta de dente	Norte	859	31	26629

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

- ▶ Qual o elemento de índice 2 e coluna Produto?
- ▶ Quantas maioneses foram vendidas?
- ▶ Consegue ler a data? O que tem de diferente?

Construindo
o **conceito**

Pandas – Data Frame

1. Qual é a receita total gerada pela venda de shampoo na região Sul?
2. Quantas unidades de pasta de dente foram vendidas no total?
3. Qual é o valor unitário médio dos produtos vendidos?
4. Em que data ocorreu a venda com a maior receita e qual foi essa receita?
5. Qual produto teve a maior quantidade vendida em uma única transação?

Construindo
o **conceito**

Pandas – Data Frame

6. Qual é a receita total gerada na região Norte?
7. Qual é a diferença percentual entre a quantidade de shampoo vendido nas regiões Leste e Sul?
8. Qual é o produto menos vendido em termos de quantidade?
9. Qual é a receita média por venda para cada produto?
10. Qual região apresentou a maior receita total e qual foi essa receita?



Vamos
fazer um
quiz

**Qual é o tipo de objeto primário para
armazenar dados tabulares no Pandas?**

Data Frame

Series

List

Dictionary



Vamos
fazer um
quiz

O que são dados tabulares?

Dados organizados em linhas e colunas, como em uma tabela

Dados não estruturados, como texto livre

Dados em formato de árvore, como XML

Dados em formato de grafo, como redes sociais



Vamos
fazer um
quiz

Qual biblioteca do Python se assemelha a planilhas eletrônicas para fazer análise de dados?

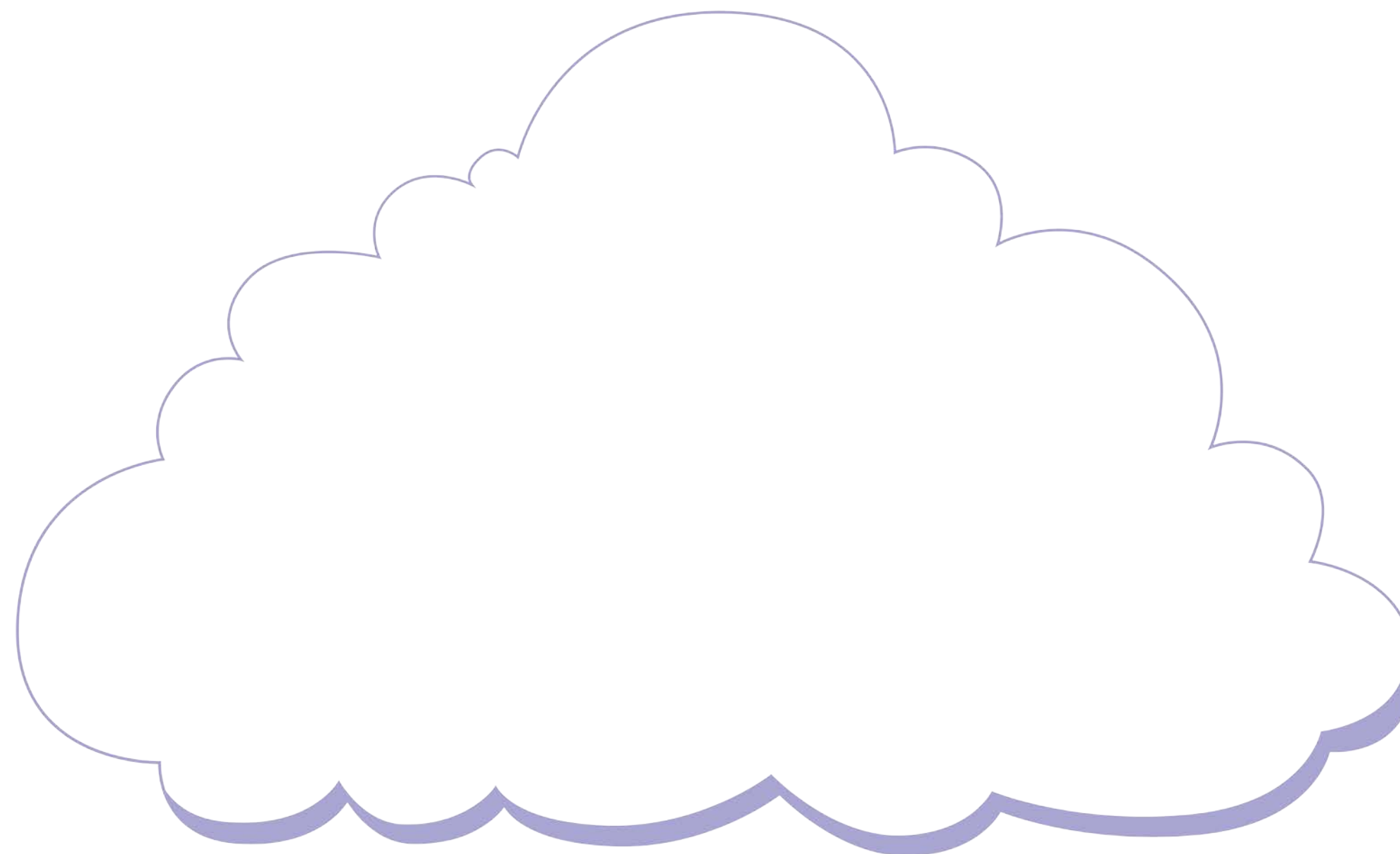
NumPy

Pandas

Scikit-learn

SciPy

Nuvem de palavras



© Getty Images

O que nós
**aprendemos
hoje?**



© Getty Images

O que nós
**aprendemos
hoje?**

Então ficamos assim...

- 1** Data Frame é um objeto bidimensional que se assemelha a planilhas eletrônicas;
- 2** Series é um objeto unidimensional que pode ser comparado a uma coluna do Data Frame;
- 3** Os índices de um Data Frame e de uma Series começam em zero.

Saiba mais

Quer dominar a análise de dados com Python?
Aprenda tudo sobre Pandas com este curso da Alura.
ALURA. Pandas: conhecendo a biblioteca.

Disponível em:
<https://cursos.alura.com.br/course/pandas-conhecendo-biblioteca>.

Acesso em: 28 jun. 2024.

Referências da aula

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

PANDAS. *Ferramenta de análise e manipulação de dados*, 2024. Disponível em: <https://pandas.pydata.org/>. Acesso em: 28 jun. 2024.

Identidade visual: imagens © Getty Images.

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**