

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**

Bibliotecas: Pandas, NumPy, SciPy, Matplotlib e Seaborn

Pandas – Transformando Data Frame

Aula 2

Código da aula: [DADOS]ANO1C2B4S25A2

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

Mapa da Unidade 5 Componente 3

Você está aqui!
Pandas –
Transformando
Data Frame

semana

25

Pandas – Combinando
Data Frame

semana

26

semana

23

Pandas – Acesso
e seleção

semana

24

Pandas – Ler e
escrever

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

Mapa da Unidade 5 Componente 3

Você está aqui!

Pandas – Transformando
Data Frame

Aula 2

Código da aula:
[DADOS]ANO1C2B4S25A2

25



Objetivos da aula

- Conhecer as operações aritméticas com as colunas de um Data Frame com a biblioteca Pandas do Python.



Recursos didáticos

- Recursos audiovisuais para exibição de vídeos e imagens;
- Acesso ao laboratório de informática e/ou 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.

Construindo o **conceito**

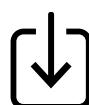
Lendo o arquivo

Abra o arquivo **notas_CD.csv** com o Pandas:

```
1 # importar a biblioteca
2 import pandas as pd
3
4 # Ler o arquivo e salvar no dataframe df
5 df = pd.read_csv("notas_CD.csv")
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Observação: o arquivo deve estar na mesma pasta na qual você está abrindo o notebook.ipynb .



Baixe o material de apoio.

Construindo
o **conceito**

Operações no Data Frame

Imagine a seguinte situação: todos os alunos receberam 1 ponto a mais no componente C1. Como você faria?

	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	Turma	Nasc	Escola
0	Micael Miguel Ehenvic Puzuho de Oliveira	9.0	7.5	9.5	19.0	2.500	Turma_5	30/01/2007	Escola 2
1	Janaína Mupru Froiri	9.0	4.0	9.0	20.0	1.250	Turma_3	18/09/2008	Escola 1
2	Álvaro Savo da Silva	9.0	8.0	7.0	15.0	1.875	Turma_4	06/07/2006	Escola 1
3	Marlon Rosatto Neto	6.0	5.5	10.0	16.0	0.125	Turma_1	16/10/2006	Escola 2
4	Sandra Inetirn	3.0	3.5	6.5	15.0	1.875	Turma_3	08/07/2006	Escola 1

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o **conceito**

Operações aritméticas

As operações aritméticas em Data Frames alinham nos rótulos de linhas e colunas.

- ▶ Isso significa que você pode **adicionar, subtrair, multiplicar** ou **dividir** Data Frames diretamente.

```
df['resultado'] = df['col1'] - df['col2']
```

```
df['resultado'] = df['col1'] + df['col2']
```

```
df['resultado'] = df['col1'] / df['col2']
```

```
df['resultado'] = df['col1'] * df['col2']
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo
o **conceito**

Operações aritméticas

Usando o operador de atribuição

Para adicionar uma coluna chamada “nova_coluna” com o valor constante de 5 para cada linha, faça o seguinte:

```
df['nova_coluna'] = 5
```

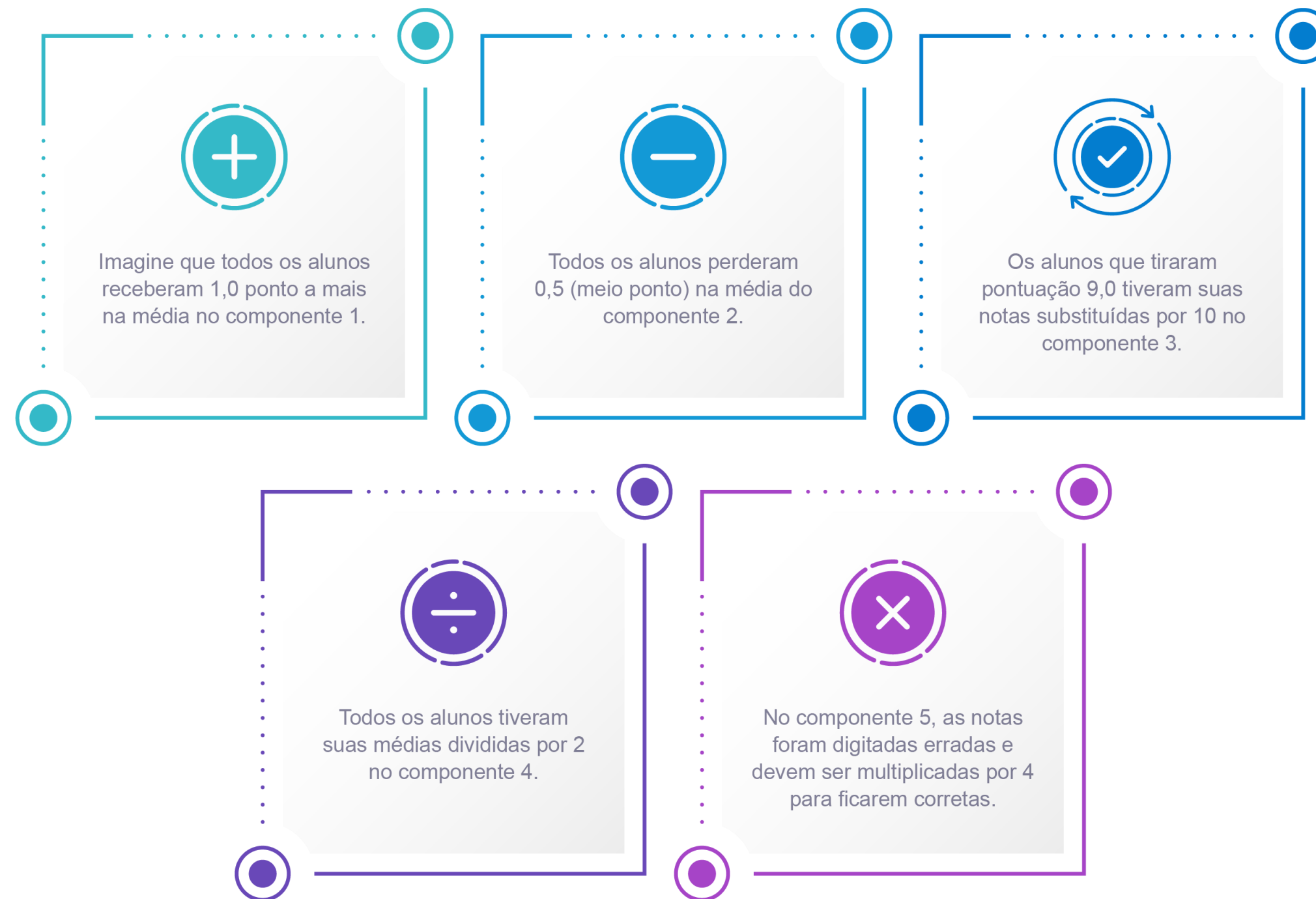
Adicionando várias colunas com o mesmo valor constante: se você quiser adicionar três novas colunas (“coluna1”, “coluna2” e “coluna3”) com o mesmo valor constante de 5, use:

```
df[['coluna1', 'coluna2', 'coluna3']] = 5
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Operações aritméticas – Situação



Elaborado especialmente para o curso com imagens © Getty Images.

Construindo o conceito

Operações aritméticas

Veja como ficará, na prática, em nosso Data Frame:

```
1 # Somar 1 ponto na nota de C1
2 df['C1'] = df['C1 - Introdução a Ciência de Dados, Ferramentas e Jargões da Área'] + 1
```

```
1 # Tirar meio ponto de todos os alunos na C2
2
3 df['C2'] = df['C2 - Ética e Responsabilidade em Inteligência Artificial'] - 0.5
```

```
1 # Substituir a nota 9 por 10 na C3
2
3 df['C3'] = df['C3 - Programação Aplicada a Ciência de Dados'].replace(9,10)
```

```
1 # Dividir a nota da C4 por 2
2
3 df['C4'] = df['C4 - Lógica, Algoritmo e Operação de Planilhas Eletrônicas'] / 2
```

```
1 # Dividir a nota da C5 por 4
2
3 df['C5'] = df['C5 - Banco de Dados e Computação em Nuvem'] * 4
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo
o **conceito**

Manipular Data Frame

No nosso Data Frame, faça **df.describe** e interprete os resultados.

```
1 df.describe()
```

	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	C1	C2	C3	C4	C5
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000
mean	6.705000	7.260000	8.320000	15.840000	1.448750	7.705000	6.760000	8.450000	7.920000	5.795000
std	2.490482	2.996193	1.575571	3.454524	0.952153	2.490482	2.996193	1.665909	1.727262	3.808613
min	0.000000	0.500000	4.000000	8.000000	0.000000	1.000000	0.000000	4.000000	4.000000	0.000000
25%	5.000000	5.000000	7.500000	13.000000	0.500000	6.000000	4.500000	7.500000	6.500000	2.000000
50%	7.250000	8.500000	8.750000	16.000000	1.625000	8.250000	8.000000	9.000000	8.000000	6.500000
75%	9.000000	10.000000	9.500000	19.250000	2.500000	10.000000	9.500000	10.000000	9.625000	10.000000
max	9.000000	10.000000	10.000000	20.000000	2.500000	10.000000	9.500000	10.000000	10.000000	10.000000

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo
o **conceito**

Manipulação de Data Frame

Como calcular a **média** entre as componentes C1, C2, C3, C4 e C5?

```
1 df['media'] = (df['C1'] + df['C2'] + df['C3'] + df['C4'])/4
```

```
1 df
```

	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	Turma	Nasc	Escola	C1	C2	C3	C4	C5	media
0	Micael Miguel Ehenvic Puzuho de Oliveira	9.0	7.5	9.5	19.0	2.500	Turma_5	30/01/2007	Escola_2	10.0	7.0	9.5	9.5	10.0	9.000
1	Janaína Mupru Froiri	9.0	4.0	9.0	20.0	1.250	Turma_3	18/09/2008	Escola_1	10.0	3.5	10.0	10.0	5.0	8.375
2	Álvaro Savo da Silva	9.0	8.0	7.0	15.0	1.875	Turma_4	06/07/2006	Escola_1	10.0	7.5	7.0	7.5	7.5	8.000
3	Marlon Rosatto Neto	6.0	5.5	10.0	16.0	0.125	Turma_1	16/10/2006	Escola_2	7.0	5.0	10.0	8.0	0.5	7.500
4	Sandra Inetirn	3.0	3.5	6.5	15.0	1.875	Turma_3	08/07/2006	Escola_1	4.0	3.0	6.5	7.5	7.5	5.250

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo
o **conceito**

Manipulação de Data Frame

E se fosse **média ponderada**? Ou seja, C1 tem peso 2, C2 tem peso 3, C3 peso 3, C4 e C5 tem peso 1.

```
|: 1 df['media_ponderada'] = (2* df['C1'] + 3* df['C2'] + 3* df['C3'] + df['C4'] + df['C5'])/10
```

```
|: 1 df
```

	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	Turma	Nasc	Escola	C1	C2	C3	C4	C5	media	media_ponderada
0	Micael Miguel Ehenvic Puzuho de Oliveira	9.0	7.5	9.5	19.0	2.500	Turma_5	30/01/2007	Escola_2	10.0	7.0	9.5	9.5	10.0	9.000	8.85
1	Janaína Mupru Froiri	9.0	4.0	9.0	20.0	1.250	Turma_3	18/09/2008	Escola_1	10.0	3.5	10.0	10.0	5.0	8.375	8.05
2	Álvaro Savo da Silva	9.0	8.0	7.0	15.0	1.875	Turma_4	06/07/2006	Escola_1	10.0	7.5	7.0	7.5	7.5	8.000	7.85
3	Marlon Rosatto Neto	6.0	5.5	10.0	16.0	0.125	Turma_1	16/10/2006	Escola_2	7.0	5.0	10.0	8.0	0.5	7.500	7.50
4	Sandra Inetirn	3.0	3.5	6.5	15.0	1.875	Turma_3	08/07/2006	Escola_1	4.0	3.0	6.5	7.5	7.5	5.250	5.15

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



Vamos
fazer um
quiz

Qual método do Pandas fornece informações sobre as colunas, tipos de dados e quantidade de valores não nulos em um Data Frame?

`df.describe()`

`df.info()`

`df.head()`

`df.shape`



Vamos
fazer um
quiz

Qual método é usado para contar a quantidade de ocorrências de cada valor em uma coluna do Data Frame?

`df['coluna'].count_valu
es()`

`df['coluna'].value_co
unts()`

`df['coluna'].unique()`

`df['coluna'].count_uni
que()`



Vamos
fazer um
quiz

Para selecionar apenas as colunas “idade” e “gênero” de um Data Frame chamado df, qual código você usaria?

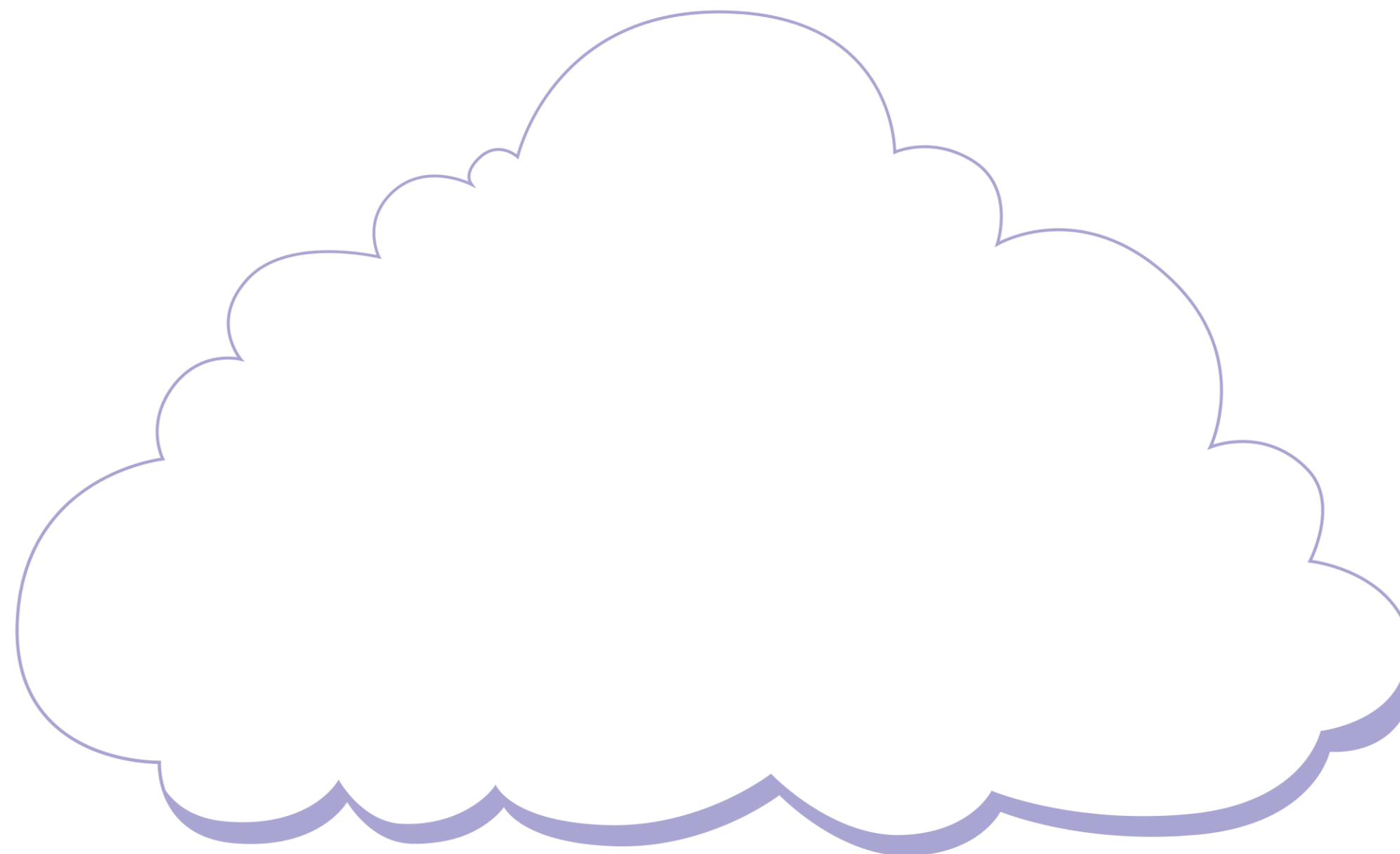
```
df[['idade', 'gênero']]
```

```
df['idade', 'gênero']
```

```
df.select_columns(['i  
dade', 'gênero'])
```

```
df.get_columns(['ida  
de', 'gênero'])
```

Nuvem de palavras



© Getty Images

O que nós
**aprendemos
hoje?**



© Getty Images

O que nós
**aprendemos
hoje?**

Então, ficamos assim:

- 1** Verificamos que as operações aritméticas funcionam bem com as colunas do Data Frame;
- 2** Podemos atribuir uma constante a uma coluna apenas usando o símbolo de atribuição;
- 3** Vimos, na prática, como os métodos atuam a partir de um exemplo com notas de alunos.

Saiba mais

Que tal encarar o desafio de aprender Pandas em dez minutos? Acesse o guia abaixo e traduza para português para você saber tudo sobre essa biblioteca!

PANDAS. *10 minutes to pandas*, [s.d.]. Disponível em: https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html#viewing-data. Acesso em: 11 jul. 2024.

Referências da aula

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

PANDAS. *Pandas documentation*, 10 abr. 2024. Disponível em: <https://pandas.pydata.org/docs/>. Acesso em: 11 jul. 2024.

Identidade visual: imagens © Getty Images

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**