

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**

Bibliotecas: Pandas, NumPy, SciPy, Matplotlib e Seaborn

Pandas: acesso e seleção

Aula 4

Código da aula: [DADOS]ANO1C2B3S23A4

**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

23

Aula 4

Código da aula:
[DADOS]ANO1C2B3S23A4



Objetivos da aula

- Aprender a selecionar e acessar o Data Frame.



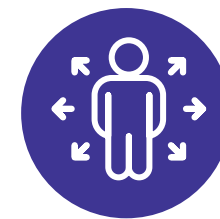
Recursos didáticos

- 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.

Construindo o conceito

Pandas – Acessando

Para acessar os dados em um Data Frame no Pandas, você pode usar vários métodos. Aqui estão alguns dos mais comuns:

- **Selecionar uma coluna específica:** use `df['nome_da_coluna']` para acessar os dados de uma coluna específica.

```
import pandas as pd

# Criando um DataFrame de exemplo
df = pd.DataFrame({
    'A': [1, 2, 3],
    'B': [4, 5, 6],
    'C': [7, 8, 9]
}, index=['linha1', 'linha2', 'linha3'])
```

```
# Acessando a coluna 'A'
df['A']
```

```
linha1    1
linha2    2
linha3    3
Name: A, dtype: int64
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Pandas – Acessando

- ▶ **Selecionar várias colunas:** Use `df[['nome_da_coluna1', 'nome_da_coluna2']]` para acessar múltiplas colunas.
- ▶ **Acessar uma linha pelo índice:** use `df.iloc[indice]` para acessar uma linha pelo seu índice numérico.

```
# Acessando as colunas 'A' e 'C'  
df[['A', 'C']]
```

	A	C
linha1	1	7
linha2	2	8
linha3	3	9

```
# Acessando a linha com índice 0  
df.iloc[0]
```

```
A    1  
B    4  
C    7  
Name: linha1, dtype: int64
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Pandas – Acessando

- ▶ **Acessar uma linha pelo rótulo do índice:** use `df.loc['rotulo']` para acessar uma linha pelo rótulo do índice.
- ▶ **Acessar um valor específico:** use `df.at[linha, 'coluna']` para acessar um único valor pela etiqueta do índice e nome da coluna.
- ▶ **Acessar um valor específico pelo índice numérico:** use `df.iat[linha, coluna]` para acessar um valor pelo índice numérico da linha e da coluna.

```
# Acessando a linha com rótulo 'linha2'  
df.loc['linha2']
```

```
A    2  
B    5  
C    8  
Name: linha2, dtype: int64
```

```
# Acessando o valor na primeira linha e coluna 'B'  
df.at['linha1', 'B']
```

```
4
```

```
# Acessando o valor na posição [2, 2] (terceira linha, terceira coluna)  
df.iat[2, 2]
```

```
9
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Colocando
em **prática**

Filmes

Com o Data Frame df_filmes abaixo, responda às perguntas usando o Pandas:

	ID	Título	Ano	Diretor	Gênero	Duração (min)
0	1	O Poderoso Chefão	1972	Francis Ford Coppola	Crime, Drama	175
1	2	A Lista de Schindler	1993	Steven Spielberg	História, Drama	195
2	3	Forrest Gump	1994	Robert Zemeckis	Romance, Drama	142
3	4	O Senhor dos Anéis	2001	Peter Jackson	Aventura, Fantasia	178
4	5	O Silêncio dos Inocentes	1991	Jonathan Demme	Crime, Drama	118



Próxima aula



Em grupo

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Colocando
em **prática**

Filmes

1. Como você acessaria a coluna 'Título' do Data Frame `df_filmes`?
2. Se você quisesse ver apenas os filmes lançados após o ano 2000, que código você usaria?
3. Como você selecionaria o filme *O senhor dos anéis* e todas as suas informações no Data Frame?
4. Qual comando você usaria para obter o número total de gols marcados pelos jogadores listados no Data Frame `df_futebol`?
5. Se você precisasse acessar a duração do filme *Forrest Gump*, que método do Pandas você utilizaria?



Próxima aula



Em grupo

Ser
sempre +

Situação

Você, como analista de dados em uma grande empresa de varejo, recebeu a tarefa de analisar as vendas dos últimos cinco anos. Os dados, contidos em vários arquivos Excel, são extensos, com milhares de linhas, contendo informações sobre vendas, clientes, produtos e datas.

Por conta das dimensões dos dados e da necessidade de manipulação e análise complexas, usar o Excel seria inviável. É aqui que entra o Pandas, uma biblioteca Python especializada em manipulação e análise de dados.

Situação fictícia elaborada especialmente para o curso.

Com o Pandas, você pode:

- carregar múltiplos arquivos Excel de forma eficiente;
- limpar e preparar os dados, removendo valores ausentes ou duplicados e convertendo tipos de dados;
- combinar dados de diferentes fontes usando operações como merge e concat;
- agrupar dados para análise sumária, como somar vendas por produto ou por região;
- realizar cálculos complexos e transformações de dados com facilidade;
- exportar os resultados da análise para novos arquivos Excel ou formatos diferentes como CSV ou JSON.

Como você convenceria o seu gestor a adotar o Python em vez do Excel?



© Getty Images

O que nós
aprendemos
hoje?

Então ficamos assim...

- 1** Aprendemos a selecionar uma coluna no Pandas utilizando a sintaxe `df['coluna']` ou `df[['col1', 'col2']]` para selecionar múltiplas colunas;
- 2** Para acessar linhas no Pandas, podemos usar `df.iloc[indice]` para índice numérico ou `df.loc['rotulo']` para rótulo do índice;
- 3** Por fim, para acessar valores, utilizamos `df.at[linha, 'coluna']` para valor por rótulo ou `df.iat[linha, coluna]` para valor por índice numérico.

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**