

**Educação  
Profissional  
Paulista**

Técnico em  
**Ciência de  
Dados**

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

## **Pandas – Ler e escrever**

Aula 4

Código da aula: [DADOS]ANO1C2B3S24A4

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

## Mapa da Unidade 5 Componente 3

semana  
**22**  
Numpy – Manipulação  
de array

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**

**24**

### **Aula 4**

Código da aula:  
[DADOS]ANO1C2B3S24A4



## Objetivos da aula

- Praticar o conceito de ler e escrever 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.





Vamos  
fazer um  
**quiz**

**Qual método do Pandas é usado  
para ler dados de um arquivo CSV  
em um DataFrame?**

`pd.read_csv()`

`pd.read_excel()`

`pd.read_json()`

`pd.read_html()`



Vamos  
fazer um  
**quiz**

**Qual argumento é usado para  
especificar o separador “;”, por  
exemplo?**

index

encoding

sep

filepath\_or\_buffer



Vamos  
fazer um  
**quiz**

**Qual método é usado para salvar  
um DataFrame em um arquivo  
Excel (formato XLSX)?**

`df.to_csv()`

`df.to_excel()`

`df.to_json()`

`df.to_html()`





Colocando  
em **prática**

## Salvando as praias

Com o dicionário a seguir:

```
1 praias = {  
2     "Nome": ["Copacabana", "Ipanema", "Praia do Forte", "Jericoacoara"],  
3     "Localização": ["Rio de Janeiro - Brasil", "Rio de Janeiro - Brasil", "Bahia - Brasil", "Ceará - Brasil"],  
4     "Tipo": ["Urbana", "Urbana", "Natural", "Natural"],  
5     "Atividades": ["Banho de mar e caminhada", "Surfe e futevolei", "Observação de tartarugas", "Kitesurf e windsurf"]  
6 }
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Crie um dataframe e use:

- o nome praia para salvar em arquivo csv, sem índice.
- o nome praia\_br para salvar em arquivo csv, separado por “;”.
- o nome praias\_lindas para salvar em arquivo csv, com encoding “latin1”.



**arquivo: ipynb**



**Durante a aula**



**Em grupo**



## Colocando em **prática**

# Abrindo as praias

1. Faça a leitura de um arquivo csv diretamente da URL:  
GITHUB. Disponível em:  
<https://raw.githubusercontent.com/remarchese/seduc/main/arquivos/praias.csv>. Acesso em: 3 jul. 2024.

2. Deixe seu dataframe assim:

	Nome	Localização	Tipo	Atividades
0	Copacabana	Rio de Janeiro - Brasil	Urbana	Banho de mar e caminhada
1	Ipanema	Rio de Janeiro - Brasil	Urbana	Surfe e futevolei
2	Praia do Forte	Bahia - Brasil	Natural	Observação de tartarugas
3	Jericoacoara	Ceará - Brasil	Natural	Kitesurf e windsurf

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



**arquivo: ipynb**



**Durante a aula**



**Em grupo**

Ser  
**sempre +**

## Situação

Carlos é um estudante do curso técnico de Ciência de Dados e está aprendendo a manipular dados em Python. Ele já está familiarizado com a biblioteca Pandas e sabe ler e processar dados de arquivos CSV. No entanto, navegando pela internet, ele encontrou um arquivo no formato Parquet.

Ele percebeu que o mesmo arquivo, em formato diferente, tinha um tamanho muito menor em Parquet do que em csv ou xlsx.

Carlos nunca tinha visto um arquivo Parquet antes, mas está confiante em suas habilidades com o Pandas. Ele sabe que o Parquet é um formato de armazenamento colunar eficiente para grandes conjuntos de dados, porém, não tem certeza de como começar a trabalhar com ele.

- ▶ Como ajudar o Carlos a abrir um arquivo Parquet?
- ▶ Qual é a estratégia usada para encontrar uma solução para abrir esse tipo de arquivo?

Situação fictícia elaborada especialmente para o curso.



# Então ficamos assim...

- 1 Praticamos salvar e abrir arquivos com a biblioteca Pandas o dataframe.

O que nós  
**aprendemos  
hoje?**

© Getty Images



# Saiba mais

Está com dificuldades para remover linhas e colunas indesejadas no Pandas? Desvende os segredos da limpeza de dados com esta técnica essencial!

MATHEUS, Y. *Como remover linhas e colunas no Pandas*. Alura, 25 dez. 2019. Disponível em: <https://www.alura.com.br/artigos/como-remover-linhas-e-colunas-no-pandas>. 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**