

Educação Profissional Paulista

Técnico em
**Ciência de
Dados**

Secretaria da
Educação



SÃO PAULO
GOVERNO DO ESTADO

Bibliotecas: Pandas, NumPy, SciPy, Matplotlib e Seaborn

Pandas – Manipulação de dados

Aula 4

Código da aula: [DADOS]ANO1C2B4S27A4

Mapa da Unidade 5 Componente 3

Bibliotecas: Pandas,
NumPy, SciPy,
Matplotlib e
Seaborn

semana
23

Pandas: Acesso e
seleção

semana
25

Pandas:
Transformando data
frame

semana
27

Você está aqui!
Pandas:
Manipulação de
dados

Bibliotecas: Pandas,
NumPy, SciPy,
Matplotlib e
Seaborn

Mapa da Unidade 5 Componente 3

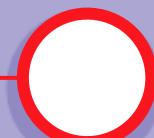
Você está aqui!

Pandas: Manipulação de datas

Aula 4

Código da aula: [DADOS]ANO1C2B4S27A4

27





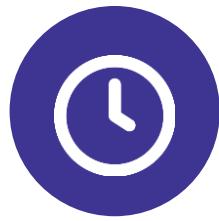
Objetivos da Aula

- Aprender sobre como é a prática de datas no Pandas.



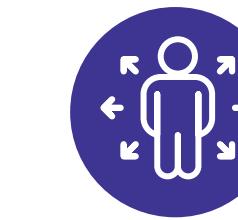
Recursos Didáticos

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



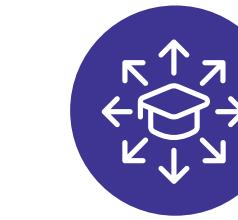
Duração da Aula

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.

Datas

No Pandas, as datas são tratadas por meio do tipo de dados **datetime64**. Isso permite que você trabalhe facilmente com datas e horas em seus conjuntos de dados.

Conheça algumas funcionalidades básicas relacionadas a datas no Pandas:

1. Criando Datas:

Você pode criar datas usando o construtor **pd.to_datetime()**, passando uma *string* no formato de data desejado.

```
import pandas as pd
date = pd.to_datetime('2024-06-02')
date
```

```
Timestamp('2024-06-02 00:00:00')
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Datas

2. Indexação com Datas:

Você pode usar datas como índices em um DataFrame.

```
: data = pd.DataFrame(  
    data=[1, 2, 3], index=pd.to_datetime(["2024-06-01", "2024-06-02", "2024-06-03"]))  
:  
: data  
:      0  
: 2024-06-01  1  
: 2024-06-02  2  
: 2024-06-03  3
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Datas

3. Filtragem por Data:

Você pode filtrar dados com base em datas.

```
data_filtrada = data['2024-06-02':'2024-06-03']  
data_filtrada
```

	0
2024-06-02	2
2024-06-03	3

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Datas

4. Resampling (reamostragem):

Você pode alterar a frequência dos dados usando o método resample().

```
data_mes = data.resample('ME').sum()  
data_mes
```

	0
2024-06-30	6

```
data_diaria = data.resample('D').sum()  
data_diaria
```

	0
2024-06-01	1
2024-06-02	2
2024-06-03	3

Nesse caso, há a ação de reamostragem por mês e somamos os valores.
No segundo caso, reamostragem por dia.

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Datas

5. Operações com Datas:

Você pode realizar operações matemáticas com datas, como adição e subtração.

```
hoje = '2024-06-02'
```

```
hoje
```

```
'2024-06-02'
```

```
amanha = date + pd.Timedelta(days=1)
```

```
amanha
```

```
Timestamp('2024-06-03 00:00:00')
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Datas

6. Diferença entre Datas:

Você pode calcular a diferença entre duas datas.

```
date2 = pd.to_datetime('2024-06-02')
date1 = pd.to_datetime('2024-03-20')
```

```
diferenca = date2 - date1
diferenca
```

```
Timedelta('74 days 00:00:00')
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Colocando
em **prática**

Transformando data

Turma, após a explicação do conteúdo, agora é o momento de vocês colocarem em prática o aprendizado de formatação internacional. Realizem o proposto:

Transformem a data "02/06/2024" em "2024-06-02"



Próxima semana



**Em grupos de até
quatro alunos**

Dicas:

- Pesquisem nas anotações e conteúdos das aulas;
- Conversem entre si e troquem opiniões;
- Realizem testes antes de enviar a resposta final.



© Getty Images

O que nós
aprendemos
hoje?

Então ficamos assim...

- 1** O método `pd.to_datetime()` permite criar objetos de data e hora a partir de *strings* em diversos formatos, facilitando a manipulação de datas em análises de dados.
- 2** Datas podem ser índices em DataFrames, permitindo indexação e filtragem eficientes. A filtragem por data é intuitiva, utilizando intervalos ou datas específicas para selecionar os dados desejados.
- 3** A função `resample()` altera a frequência dos dados para diferentes granularidades temporais, útil para agregação e análise. Operações com datas, como adição e subtração usando `pd.Timedelta()`, facilitam cálculos temporais.



Ser
sempre +

Situação

Tema: Transformando data em formato internacional

Havia um estagiário dedicado chamado Pedro, que trabalhava em uma empresa multinacional. Pedro era habilidoso em muitas áreas, mas quando se tratava de manipular datas no sistema internacional, ele se sentia inseguro.

Certa vez, Pedro recebeu uma tarefa urgente que exigia inserir a data de 02/06/2024 no sistema da empresa, mas ele não sabia como transformá-la no formato internacional padrão: 2024-06-02. Ele tentou várias vezes por meio de consultas on-line, porém a data sempre ficava no formato incorreto.

Mesmo nervoso diante da situação, Pedro percebeu que precisava se dedicar e aprender essa habilidade para o seu trabalho. Ele se dedicou estudando e praticando a conversão de datas.

Com determinação e paciência, Pedro finalmente dominou a arte de transformar datas no formato internacional, porém o gestor estava esperando um retorno há dias, e então, ele achou importante comunicar o gestor via e-mail sobre a situação de atraso da entrega.

Situação fictícia elaborada especialmente para o curso



Ser
sempre +

Ação

Em grupos de até quatro alunos, analisem a situação de Pedro, e em seguida, respondam às questões propostas:

- ▶ Quais são os aspectos essenciais que devem ser abordados por Pedro ao escrever um e-mail de justificativa para o atraso na entrega da tarefa?

- ▶ Ao se colocar no lugar de Pedro, como seria o texto do seu e-mail a ser enviado ao gestor?



Saiba mais

Quer entender mais sobre como definir data e hora no Python?

Conheça:

ORESTES, Y. *Python datetime: como faço para definir data e hora em Python?* Alura, 18 set. 2023. Disponível em: <https://www.alura.com.br/artigos/lidando-com-datas-e-horarios-no-python/>. Acesso em: 30 jul. 2024.

Referências da aula

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

PANDAS. User Guide: merge, join, concatenate and compare, [s.d.]. Disponível em: https://pandas.pydata.org/docs/user_guide/merging.html. Acesso em: 13 ago. 2024.

Identidade visual: imagens © Getty Images.

Educação Profissional Paulista

Técnico em
**Ciência de
Dados**

Secretaria da
Educação



SÃO PAULO
GOVERNO DO ESTADO