

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 3

Código da aula: [DADOS]ANO1C2B4S27A3

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

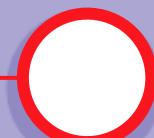
Você está aqui!

Pandas: Manipulação de datas

Aula 3

Código da aula: [DADOS]ANO1C2B4S27A3

27





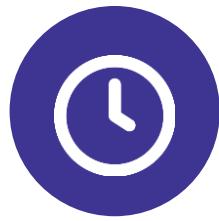
Objetivos da Aula

- Praticar os métodos de agregação do groupby do Pandas, Python.



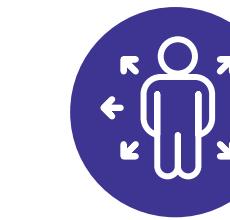
Recursos Didáticos

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



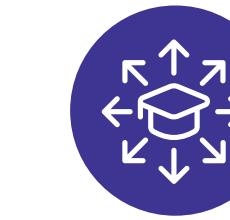
Duração da Aula

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.

Mercado de frutas

De acordo com o DataFrame abaixo, vamos descobrir na prática a agregação do *groupby*.

```
import pandas as pd

frutas_data = {
    "Fruta": ["Maçã", "Banana", "Laranja", "Uva", "Maçã", "Abacaxi",
              "Morango", "Melancia", "Maçã", "Laranja"],
    "Mercado": ["Mercado A", "Mercado B", "Mercado C", "Mercado A",
                "Mercado B", "Mercado C", "Mercado A", "Mercado B",
                "Mercado C", "Mercado A"],
    "Quantidade": [10, 15, 8, 12, 20, 14, 18, 25, 9, 11]
}

df_frutas = pd.DataFrame(frutas_data)
df_frutas
```

	Fruta	Mercado	Quantidade
0	Maçã	Mercado A	10
1	Banana	Mercado B	15
2	Laranja	Mercado C	8
3	Uva	Mercado A	12
4	Maçã	Mercado B	20
5	Abacaxi	Mercado C	14
6	Morango	Mercado A	18
7	Melancia	Mercado B	25
8	Maçã	Mercado C	9
9	Laranja	Mercado A	11

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Mercado de frutas

- ▶ Quantas unidades de cada fruta foram vendidas no total?

```
# Quantas unidades de cada fruta foram vendidas no total?  
frutas_total = df_frutas.groupby('Fruta')['Quantidade'].sum()  
frutas_total
```

```
Fruta  
Abacaxi      14  
Banana       15  
Laranja      19  
Maçã          39  
Melancia     25  
Morango      18  
Uva           12  
Name: Quantidade, dtype: int64
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Mercado de frutas

- ▶ Qual foi o número total de frutas vendidas em cada mercado?

```
# Qual foi o número total de frutas vendidas em cada mercado?  
mercado_total = df_frutas.groupby('Mercado')['Quantidade'].sum()  
mercado_total
```

```
Mercado  
Mercado A    51  
Mercado B    60  
Mercado C    31  
Name: Quantidade, dtype: int64
```

```
# Qual foi o número total de frutas vendidas em cada mercado?  
mercado_total = df_frutas.groupby('Mercado')['Quantidade'].sum().reset_index()  
mercado_total
```

	Mercado	Quantidade
0	Mercado A	51
1	Mercado B	60
2	Mercado C	31

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Mercado de frutas

- ▶ Quantas unidades de cada fruta foram vendidas em cada mercado?
Utilizando reset_index():

```
# Quantas unidades de cada fruta foram vendidas em cada mercado?
frutas_mercado = df_frutas.groupby(['Fruta', 'Mercado'])['Quantidade'].sum().reset_index()
frutas_mercado
```

	Fruta	Mercado	Quantidade
0	Abacaxi	Mercado C	14
1	Banana	Mercado B	15
2	Laranja	Mercado A	11
3	Laranja	Mercado C	8
4	Maçã	Mercado A	10
5	Maçã	Mercado B	20
6	Maçã	Mercado C	9
7	Melancia	Mercado B	25
8	Morango	Mercado A	18
9	Uva	Mercado A	12

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Mercado de frutas

- ▶ Quantas unidades de cada fruta foram vendidas em cada mercado?

```
# Quantas unidades de cada fruta foram vendidas em cada mercado?  
frutas_mercado = df_frutas.groupby(['Fruta', 'Mercado'])['Quantidade'].sum()  
frutas_mercado
```

```
Fruta      Mercado  
Abacaxi    Mercado C    14  
Banana     Mercado B    15  
Laranja    Mercado A    11  
          Mercado C     8  
Maçã       Mercado A    10  
          Mercado B     20  
          Mercado C     9  
Melancia   Mercado B    25  
Morango    Mercado A    18  
Uva        Mercado A    12  
Name: Quantidade, dtype: int64
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Mercado de frutas

- ▶ E quantas **unidades de frutas** foram vendidas em cada mercado para **cada fruta**? Utilizando `reset_index()`:

```
# Quantas unidades de frutas foram vendidas em cada mercado para cada fruta?  
mercado_fruta = df_frutas.groupby(['Mercado', 'Fruta'])['Quantidade'].sum().reset_index()  
mercado_fruta
```

	Mercado	Fruta	Quantidade
0	Mercado A	Laranja	11
1	Mercado A	Maçã	10
2	Mercado A	Morango	18
3	Mercado A	Uva	12
4	Mercado B	Banana	15
5	Mercado B	Maçã	20
6	Mercado B	Melancia	25
7	Mercado C	Abacaxi	14
8	Mercado C	Laranja	8
9	Mercado C	Maçã	9

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Mercado de frutas

- ▶ E quantas **unidades de frutas** foram vendidas em cada mercado para **cada fruta**? Veja como é em não utilizar `reset_index()`:

```
# Quantas unidades de frutas foram vendidas em cada mercado para cada fruta?  
mercado_fruta = df_frutas.groupby(['Mercado', 'Fruta'])['Quantidade'].sum()  
mercado_fruta
```

```
Mercado      Fruta  
Mercado A    Laranja     11  
                  Maçã       10  
                  Morango    18  
                  Uva        12  
Mercado B    Banana     15  
                  Maçã       20  
                  Melancia   25  
Mercado C    Abacaxi    14  
                  Laranja    8  
                  Maçã       9  
Name: Quantidade, dtype: int64
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Mercado de frutas

- ▶ Qual é a **média** de unidades de frutas que foram vendidas em cada mercado para cada fruta?

```
mercado_fruta = df_frutas.groupby(['Mercado', 'Fruta'])['Quantidade'].mean()  
mercado_fruta
```

```
Mercado      Fruta  
Mercado A    Laranja     11.0  
                  Maçã       10.0  
                  Morango    18.0  
                  Uva        12.0  
Mercado B    Banana     15.0  
                  Maçã       20.0  
                  Melancia   25.0  
Mercado C    Abacaxi    14.0  
                  Laranja    8.0  
                  Maçã       9.0  
Name: Quantidade, dtype: float64
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Mercado de frutas

- ▶ Vamos agora saber qual é a **maior quantidade** de frutas que foram vendidas em cada mercado para cada fruta:

```
mercado_fruta = df_frutas.groupby(['Mercado', 'Fruta'])['Quantidade'].max()  
mercado_fruta
```

Mercado	Fruta	Quantidade
Mercado A	Laranja	11
	Maçã	10
	Morango	18
	Uva	12
Mercado B	Banana	15
	Maçã	20
	Melancia	25
Mercado C	Abacaxi	14
	Laranja	8
	Maçã	9

Name: Quantidade, dtype: int64

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Mercado de frutas

- ▶ Qual é a quantidade média de frutas que foram vendidas em cada mercado?

```
mercado_fruta = df_frutas.groupby(['Mercado'])['Quantidade'].mean()  
mercado_fruta
```

```
Mercado  
Mercado A    12.750000  
Mercado B    20.000000  
Mercado C    10.333333  
Name: Quantidade, dtype: float64
```

- ▶ E qual é a quantidade mínima de frutas que foram vendidas em cada mercado?

```
mercado_fruta = df_frutas.groupby(['Mercado'])['Quantidade'].min()  
mercado_fruta
```

```
Mercado  
Mercado A    10  
Mercado B    15  
Mercado C     8  
Name: Quantidade, dtype: int64
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



Colocando
em **prática**

Quadro de Medalhas

Você é um técnico em ciência de dados trabalhando com um projeto emocionante: analisar o quadro de medalhas do atletismo em competições internacionais.

Para isso, você tem em mãos um arquivo valioso, o "**quadro_medalhas.csv**".

Sua missão é mergulhar nesses dados, explorar as informações sobre as medalhas conquistadas e responder a algumas perguntas importantes.



Durante a aula



Em grupos de até quatro alunos



Baixem o material de apoio da atividade

Situação fictícia elaborada especialmente para o curso.



© Getty Images



Colocando em **prática**

Quadro de Medalhas

Após análise dos dados, responda às perguntas a seguir junto com o material de apoio disponibilizado.



Durante a aula



Em grupos de até quatro alunos



Enviem o arquivo em .ipynb no AVA

1 Quantas medalhas cada país ganhou no total ao longo dos anos?

2 Qual foi o número total de medalhas de cada tipo (Ouro, Prata, Bronze) ao longo dos anos?

3 Quantas medalhas cada país ganhou em cada ano?

4 Quantas medalhas de cada tipo cada país ganhou em cada ano?

5 Qual é o menor valor de medalhas por ano que cada país recebeu?



© Getty Images

O que nós
aprendemos
hoje?

Então ficamos assim...

- 1** Vimos como é a aplicação de agregações em agrupamento.
- 2** Utilizamos códigos de Python em problemas reais.
- 3** Praticamos o uso de `groupby` da biblioteca Pandas do Python.



Saiba mais

Groupby: divide-aplica-combina.

Sabia que além de agregações, o groupby pode fazer transformações e filtros?

Conheça:

PANDAS. *User Guide: groupby: split-apply-combine*, [s.d.]. Disponível em:

https://pandas.pydata.org/docs/user_guide/groupby.html. 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: groupby: split-apply-combine*, [s.d.]. Disponível em: https://pandas.pydata.org/docs/user_guide/groupby.html. Acesso em: 30 jul. 2024.

Identidade visual: imagens © Getty Images.

Educação Profissional Paulista

Técnico em
**Ciência de
Dados**

Secretaria da
Educação



SÃO PAULO
GOVERNO DO ESTADO