

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**

Bibliotecas: Pandas, NumPy, SciPy, Matplotlib e Seaborn

Pandas: acesso e seleção

Aula 1

Código da aula: [DADOS]ANO1C2B3S23A1

**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

Aula 1

Código da aula: [DADOS]ANO1C2B3S23A1

23



Objetivos da aula

- Conhecer a biblioteca Pandas.



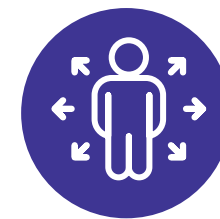
Recursos didáticos

- 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 com base 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.

Microsoft anuncia integração do Python com o Excel

A Microsoft anunciou a integração da linguagem de programação com o Excel sem necessidade de configuração adicional

Por **William Schendes**, editado por **Bruno Capozzi** | ⌚ 22/08/2023 16h29, atualizada em 23/08/2023 21h17

Primeiras
ideias

Qual é seu sentimento ao ler a reportagem?

Como você vê a importância do Python?

Como usar o Python com planilhas eletrônicas?

O que acontecerá com quem não sabe Python?

Ponto de partida

E agora?

O Excel tem um limite de **1.048.576** linhas e **16.384** colunas por planilha. Esses limites são projetados para acomodar uma vasta gama de necessidades de dados e garantir que o programa possa operar de maneira eficiente em diferentes configurações de hardware e software. Fonte: MICROSOFT, [s.d.]

Se quisermos analisar os usuários do WhatsApp em 2022, no qual tem mais de 2 bilhões de usuários no mundo, consigo fazer por planilhas eletrônicas? Não!

Fonte: DEAN, 2022.

Para conjuntos de dados que excedem esses limites, é recomendável considerar outras ferramentas, como bancos de dados ou softwares especializados em manipulação de grandes volumes de dados, por exemplo, o Python.

Mas como usar o Python para substituir planilhas eletrônicas? Listas? Dicionários?

Construindo
o **conceito**

Dados tabulares

Dados tabulares são uma forma eficiente de organizar e analisar informações em formato de tabela.

Definição e função:

- ▶ A tabulação de dados envolve transformar informações coletadas em dados que possam ser analisados. Ela é fundamental para a interpretação e sistematização dos dados.
- ▶ Contagem de respostas e codificação dos dados coletados são os principais objetivos da tabulação. Isso ajuda na padronização das informações, contribuindo para o planejamento estratégico de empresas e tomadas de decisão.

Construindo
o **conceito**

Tipos de tabulação

- ▶ **Tabulação simples:** conta o número de casos em que uma variável ocorreu.
- ▶ **Tabulação cruzada:** envolve pelo menos duas variáveis e pode incluir entrevistas com respostas múltiplas, perguntas encadeadas, perguntas abertas etc.

Construindo
o **conceito**

Passos para fazer a tabulação de dados

1

Planeje o método de pesquisa

Defina a metodologia e os elementos de pesquisa de campo.

2

Crie o questionário de pesquisa

Elabore perguntas relevantes para coletar os dados desejados.

3

Prepare o modelo para a tabulação

Organize os dados de forma adequada para a análise.

4

Faça a tabulação de dados no modelo criado

Realize a contagem e a codificação dos dados.

5

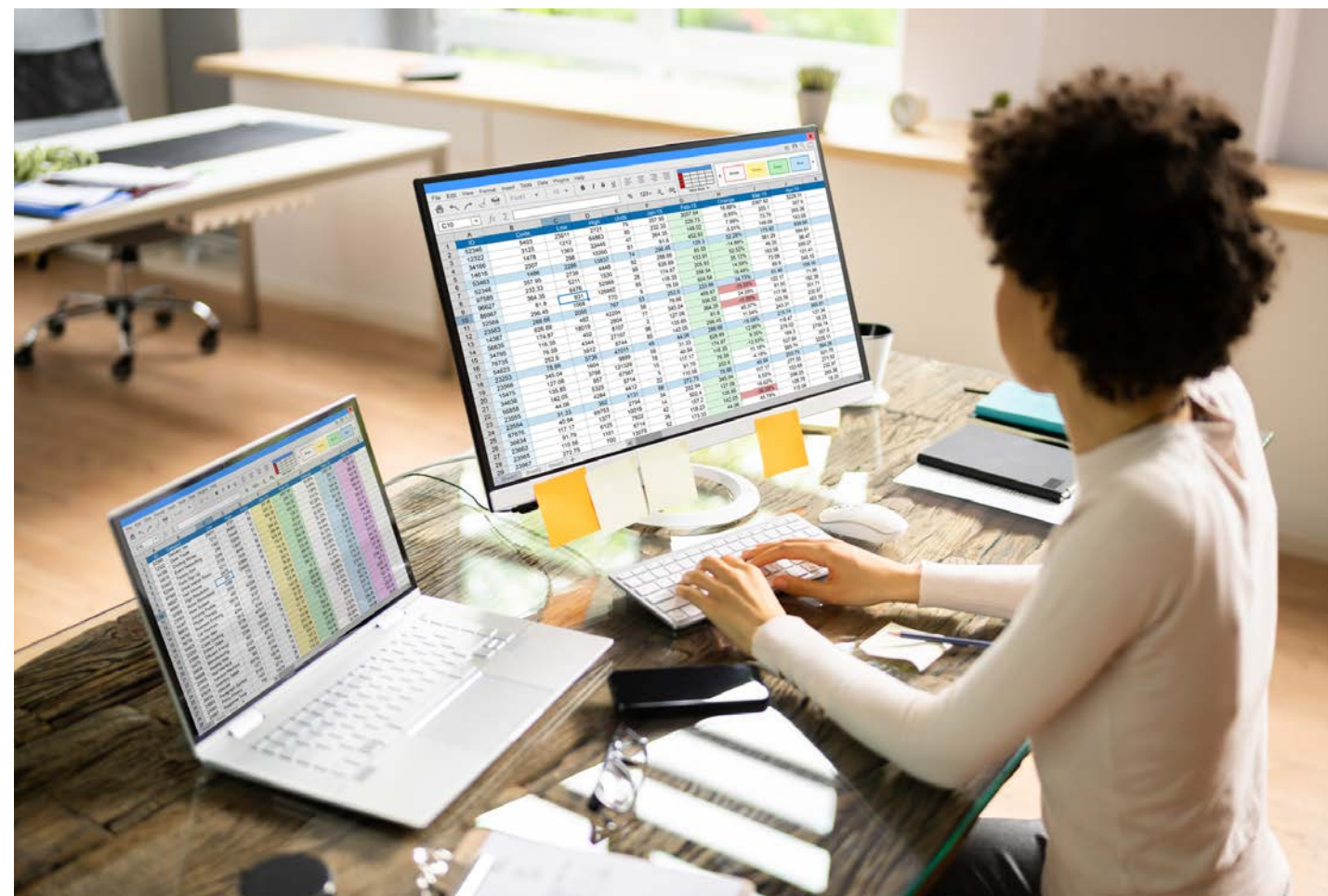
Elabore o relatório final

Sistematize os resultados obtidos.

Construindo
o **conceito**

Tecnologia e ferramentas

Existem softwares específicos para tabular os dados, além da possibilidade de criar tabelas facilmente em planilhas do Excel.



© Getty Images

Construindo
o **conceito**

Dados tabulares

No Excel, por exemplo, temos linhas e colunas:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
12														

Elaborado especialmente para o curso com a ferramenta Microsoft Excel.

Construindo
o **conceito**

Dados tabulares

E conseguimos identificar células. Por exemplo, qual é a célula da cor azul-escura? Qual cor está na célula B4?

	A	B	C	D	E
1					
2					
3					
4					
5					
6					
7					
8					

Elaborado especialmente para o curso com a ferramenta Microsoft Excel.

Construindo o **conceito**

Pandas

O **Pandas** é uma biblioteca essencial para análise de dados em Python. Ele oferece estruturas de dados e ferramentas poderosas para trabalhar com dados tabulares, como os encontrados em planilhas e bancos de dados.

O que é o Pandas?

- ▶ O Pandas é uma biblioteca de código aberto construída sobre a linguagem Python.
- ▶ Ela providencia uma abordagem rápida e flexível para trabalhar com dados relacionais (ou rotulados).

Fonte: ALMEIDA, 2023.

Construindo
o **conceito**

Pandas

- ▶ O nome “Pandas” é derivado do termo “Pan el Da ta”, relacionado ao campo de estudo da econometria.
- ▶ Essa biblioteca é amplamente utilizada em Ciência de Dados e oferece uma maneira simples e intuitiva de lidar com dados.

Construindo
o **conceito**

Pandas

Como funciona o Pandas?

O Pandas apresenta dois objetos primários:

Series: são *arrays* unidimensionais com um eixo de rótulos (ou índice) para identificar cada registro.

Data Frames: são estruturas bidimensionais que organizam dados em tabelas.

Fonte: ALMEIDA, 2023.

Construindo o **conceito**

Pandas

O Pandas é utilizado para várias atividades, incluindo:

- ▶ Limpeza e tratamento de dados;
- ▶ Análise exploratória de dados (EDA);
- ▶ Suporte em Machine Learning;
- ▶ Consultas em bancos de dados relacionais;
- ▶ Visualização de dados;
- ▶ *Webscraping* e muito mais.

Além disso, o Pandas integra-se bem com outras bibliotecas populares, como NumPy, Scikit-Learn, Seaborn, Matplotlib, entre outras.

Fonte: ALMEIDA, 2023.

Construindo
o **conceito**

Instalação do Pandas

Para instalar o Pandas, você pode usar o comando `pip install pandas` no seu ambiente Python.

```
pip install pandas
```

ou

```
conda install anaconda::pandas
```

Elaborado especialmente para o curso.



Recurso digital

Acesse a documentação do Pandas: PyData. Pandas.
Disponível em: <https://pandas.pydata.org/>. Acesso em: 28 jun. 2024.

Pandas: história

1

2008

É iniciado o desenvolvimento do Pandas.

2

2009

Pandas torna-se código aberto.

3

2012

A primeira edição do Python for Data Analysis é publicada.

4

2015

Pandas torna-se um projeto patrocinado pela NumFOCUS.

5

2018

Surge o primeiro *sprint* de desenvolvedor principal presencial.

Fonte: PANDAS, [s.d.].

Construindo
o **conceito**

Destques do Pandas

- ▶ "Um objeto **Data Frame** rápido e eficiente para manipulação de dados com indexação integrada;
- ▶ Ferramentas para **leitura** e **gravação** de dados entre estruturas de dados na memória e diferentes formatos: arquivos CSV e de texto, Microsoft Excel, bancos de dados SQL e o rápido formato HDF5;
- ▶ **Alinhamento** inteligente de dados e tratamento integrado de **dados ausentes**: obtenha alinhamento automático baseado em rótulos em cálculos e manipule facilmente dados confusos em um formato ordenado; [...]."

(PANDAS, [s.d.])

Construindo
o **conceito**

Destques do Pandas

- ▶ "**Remodelação** e dinamização flexíveis de conjuntos de dados;
- ▶ **Fatiamento** inteligente baseado em rótulos, **indexação** sofisticada e **subconjuntos** de grandes conjuntos de dados;
- ▶ As colunas podem ser inseridas e excluídas das estruturas de dados para **variar o tamanho**; [...]."

(PANDAS, [s.d.])

Construindo
o **conceito**

Destques do Pandas

- ▶ "Agregar ou transformar dados com um poderoso **mecanismo de agrupamento**, permitindo operações de divisão, aplicação e combinação em conjuntos de dados;
- ▶ **Fusão e junção** de conjuntos de dados de alto desempenho;
- ▶ A **indexação** de eixo hierárquico fornece uma maneira intuitiva de trabalhar com dados de alta dimensão em uma estrutura de dados de menor dimensão; [...]."

(PANDAS, [s.d.])

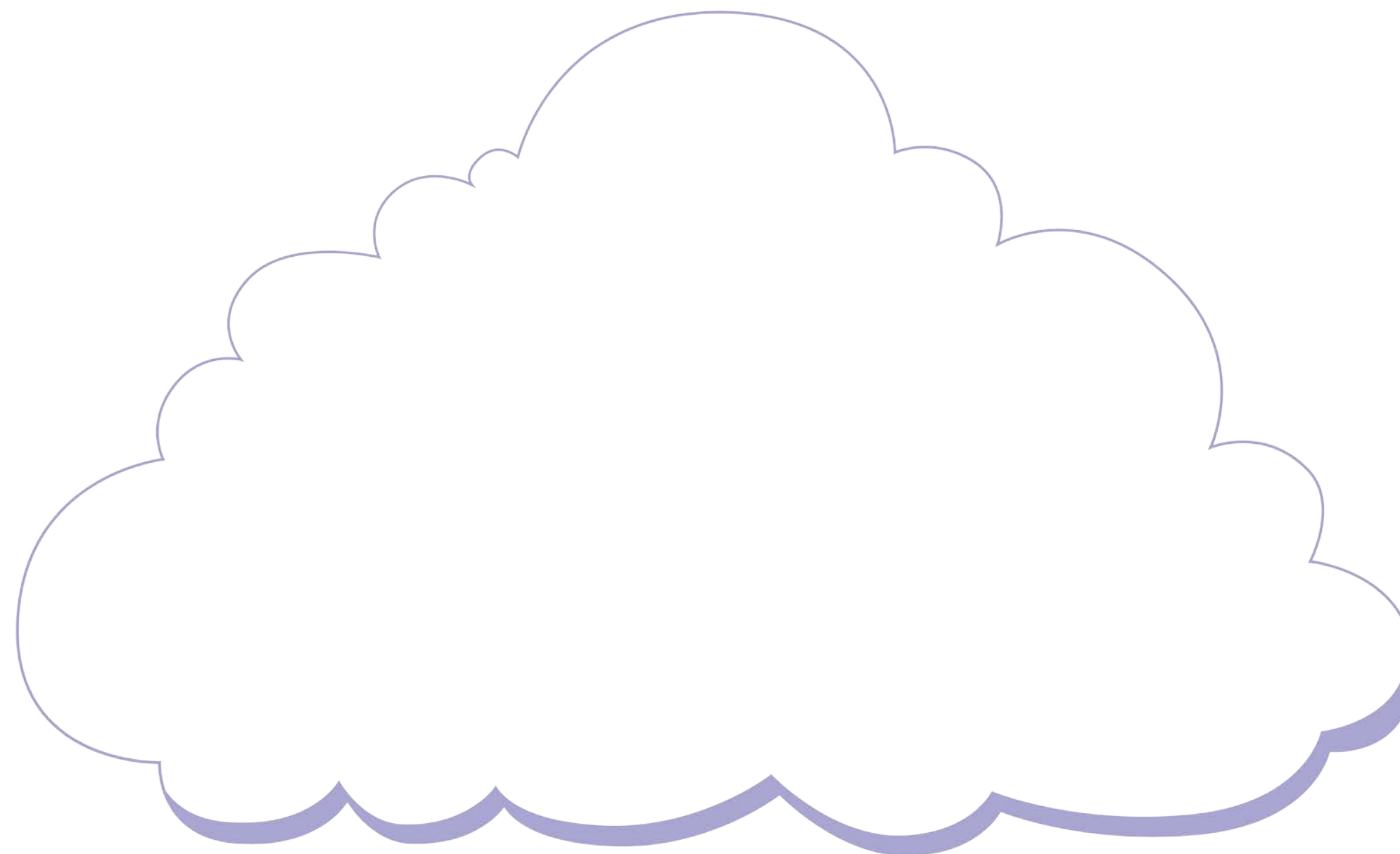
Construindo
o **conceito**

Destques do Pandas

- ▶ **"Funcionalidade de série temporal:** geração de intervalo de datas e conversão de frequência, estatísticas de janela móvel, mudança e atraso de data. Crie até mesmo compensações de horário específicas de domínio e junte séries temporais sem perder dados;
- ▶ Altamente **otimizado para desempenho**, com caminhos de código críticos escritos em Cython ou C.
- ▶ Python com Pandas está em uso em uma ampla variedade de domínios **acadêmicos e comerciais**, incluindo Finanças, Neurociências, Economia, Estatística, Publicidade, Análise da Web e muito mais."

(PANDAS, [s.d.])

Nuvem de palavras



© Getty Images

O que nós
**aprendemos
hoje?**



© Getty Images

O que nós
**aprendemos
hoje?**

Então ficamos assim...

- 1** A tabulação de dados é uma etapa crucial para transformar informações em insights úteis e apoiar processos de análise em diversos contextos.
- 2** O Pandas é uma ferramenta poderosa para manipulação e análise de dados, oferecendo funcionalidades semelhantes às do Excel e SQL.

Saiba mais

Se você quer dominar o uso do Pandas e entender como essa poderosa biblioteca pode transformar a forma como você lida com dados, não perca este artigo incrível!

ALMEIDA, M. *Pandas Python: o que é, para que serve e como instalar*. Alura, 16 ago. 2023.

Disponível em:

<https://www.alura.com.br/artigos/pandas-o-que-e-para-que-serve-como-instalar>.

Acesso em: 28 jun. 2024.

Referências da aula

DEAN, B. *Estatísticas do WhatsApp em 2022: quantas pessoas usam o aplicativo?* Semrush, 30 nov. 2022. Disponível em: <https://pt.semrush.com/blog/estatisticas-whatsapp/>. Acesso em: 28 jun. 2024.

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

MICROSOFT. *Especificações e limites do Microsoft Excel*, [s.d.]. Disponível em: <https://support.microsoft.com/pt-br/office/especifica%C3%A7%C3%B5es-e-limites-do-microsoft-excel-1672b34d-7043-467e-8e27-269d656771c3>. Acesso em: 28 jun. 2024.

PANDAS. *Ferramenta de análise e manipulação de dados*, 2024. Disponível em: <https://pandas.pydata.org/>. Acesso em: 28 jun. 2024.

PANDAS. *Sobre pandas – História do desenvolvimento*, [s.d.]. Disponível em: <https://pandas.pydata.org/about/>. Acesso em: 28 jun. 2024.

SCHENDES, W. *Microsoft anuncia integração do Python com o Excel*. Olhar Digital, 22 ago. 2023. Disponível em: https://olhardigital.com.br/2023/08/22/pro/microsoft-anuncia-integracao-do-python-com-o-excel/#google_vignette. Acesso em: 28 jun. 2024.

Identidade visual: imagens © Getty Images.

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**

Bibliotecas: Pandas, NumPy, SciPy, Matplotlib e Seaborn

Pandas: acesso e seleção

Aula 2

Código da aula: [DADOS]ANO1C2B3S23A2

**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

Aula 2

Código da aula: [DADOS]ANO1C2B3S23A2

23



Objetivos da aula

- Conhecer o conceito de Data Frame e series do Pandas.



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 com base 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

Series:

Uma Series é uma estrutura de dados unidimensional capaz de armazenar qualquer tipo de dados, como inteiros, *strings*, números flutuantes, objetos Python etc.

Cada elemento em uma Series é associado a um rótulo, também conhecido como índice.

Construindo
o **conceito**

Pandas

Data Frame:

Um Data Frame é uma estrutura de dados bidimensional, semelhante a uma tabela de banco de dados ou a uma planilha eletrônica, em que cada coluna é uma Series.

Data Frames oferecem funcionalidades ricas e flexíveis para a manipulação e análise de dados.

Construindo o conceito

Diferença entre Series e Data Frame

A Series apresenta uma estrutura unidimensional, enquanto o Data Frame tem uma estrutura bidimensional.

			Data	Produto	Região	Quantidade	Valor Unitário	Receita
0	2022-12-05							
1	2022-06-26							
2	2022-03-16	0	2022-12-05	Shampoo	Leste	360	104	37440
3	2023-06-13	1	2022-06-26	Shampoo	Sul	578	226	130628
4	2022-06-03							
	...							
3647	2023-07-21	2	2022-03-16	Pasta de dente	Sul	939	119	111741
3648	2023-11-21	3	2023-06-13	Maionese	Norte	472	83	39176
3649	2023-11-22							
3650	2023-06-06	4	2022-06-03	Pasta de dente	Norte	859	31	26629
3651	2023-02-13							

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Semelhança entre Series e Data Frame

Tanto em uma Series quanto em um Data Frame, os índices começam em zero por padrão.

			Data	Produto	Região	Quantidade	Valor Unitário	Receita
0	2022-12-05							
1	2022-06-26							
2	2022-03-16	0	2022-12-05	Shampoo	Leste	360	104	37440
3	2023-06-13	1	2022-06-26	Shampoo	Sul	578	226	130628
4	2022-06-03							
	...							
3647	2023-07-21	2	2022-03-16	Pasta de dente	Sul	939	119	111741
3648	2023-11-21	3	2023-06-13	Maionese	Norte	472	83	39176
3649	2023-11-22							
3650	2023-06-06	4	2022-06-03	Pasta de dente	Norte	859	31	26629
3651	2023-02-13							

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Pandas – Data Frame

Observe a tabela e responda:

	Data	Produto	Região	Quantidade	Valor Unitário	Receita
0	2022-12-05	Shampoo	Leste	360	104	37440
1	2022-06-26	Shampoo	Sul	578	226	130628
2	2022-03-16	Pasta de dente	Sul	939	119	111741
3	2023-06-13	Maionese	Norte	472	83	39176
4	2022-06-03	Pasta de dente	Norte	859	31	26629

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

- ▶ Qual o elemento de índice 2 e coluna Produto?
- ▶ Quantas maioneses foram vendidas?
- ▶ Consegue ler a data? O que tem de diferente?

Construindo
o **conceito**

Pandas – Data Frame

1. Qual é a receita total gerada pela venda de shampoo na região Sul?
2. Quantas unidades de pasta de dente foram vendidas no total?
3. Qual é o valor unitário médio dos produtos vendidos?
4. Em que data ocorreu a venda com a maior receita e qual foi essa receita?
5. Qual produto teve a maior quantidade vendida em uma única transação?

Construindo
o **conceito**

Pandas – Data Frame

6. Qual é a receita total gerada na região Norte?
7. Qual é a diferença percentual entre a quantidade de shampoo vendido nas regiões Leste e Sul?
8. Qual é o produto menos vendido em termos de quantidade?
9. Qual é a receita média por venda para cada produto?
10. Qual região apresentou a maior receita total e qual foi essa receita?



Vamos
fazer um
quiz

**Qual é o tipo de objeto primário para
armazenar dados tabulares no Pandas?**

Data Frame

Series

List

Dictionary



Vamos
fazer um
quiz

O que são dados tabulares?

Dados organizados em linhas e colunas, como em uma tabela

Dados não estruturados, como texto livre

Dados em formato de árvore, como XML

Dados em formato de grafo, como redes sociais



Vamos
fazer um
quiz

Qual biblioteca do Python se assemelha a planilhas eletrônicas para fazer análise de dados?

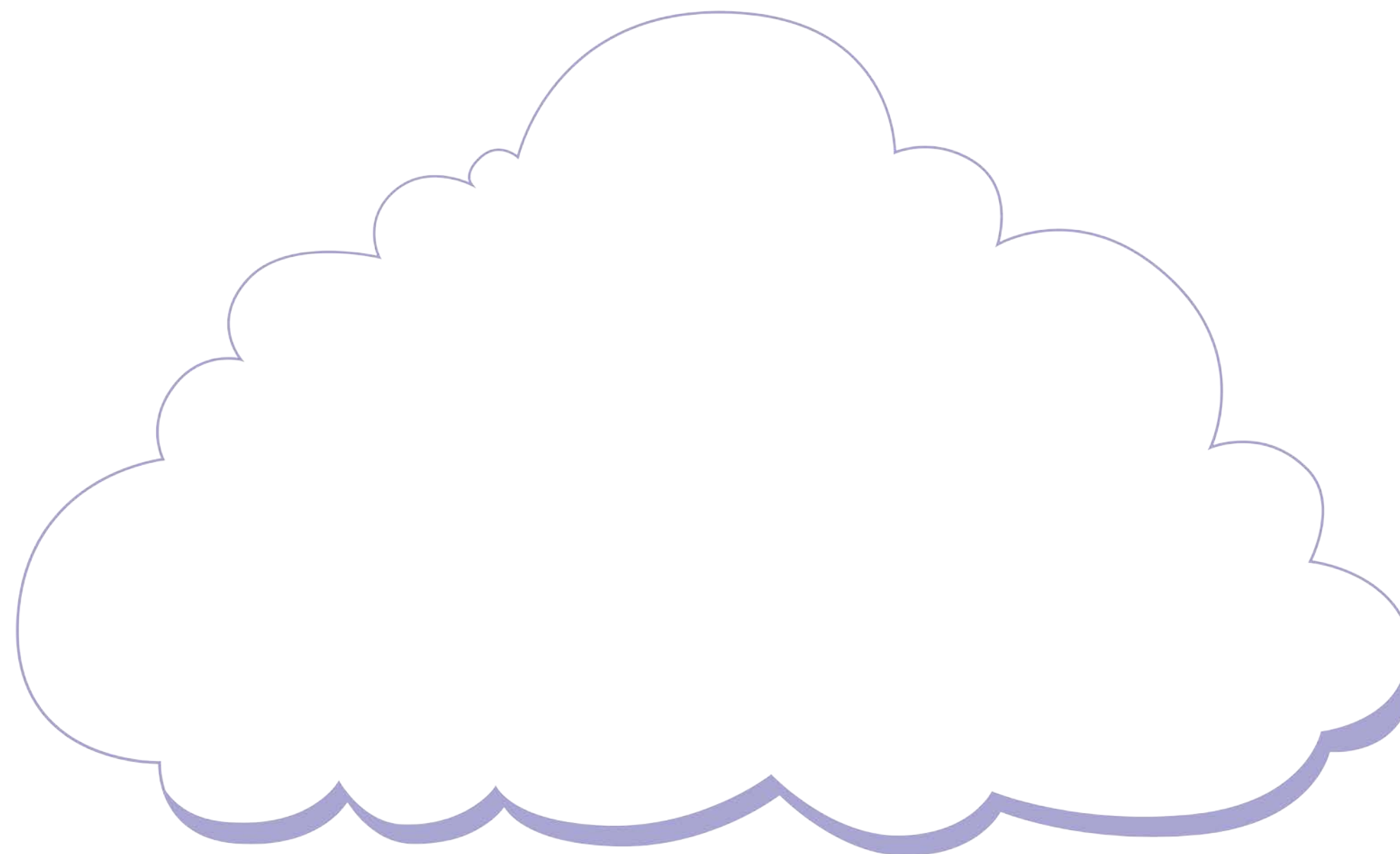
NumPy

Pandas

Scikit-learn

SciPy

Nuvem de palavras



© Getty Images

O que nós
**aprendemos
hoje?**



© Getty Images

O que nós
**aprendemos
hoje?**

Então ficamos assim...

- 1** Data Frame é um objeto bidimensional que se assemelha a planilhas eletrônicas;
- 2** Series é um objeto unidimensional que pode ser comparado a uma coluna do Data Frame;
- 3** Os índices de um Data Frame e de uma Series começam em zero.

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**

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**

Bibliotecas: Pandas, NumPy, SciPy, Matplotlib e Seaborn

Pandas: acesso e seleção

Aula 3

Código da aula: [DADOS]ANO1C2B3S23A3

**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

Aula 3

Código da aula: [DADOS]ANO1C2B3S23A3

23



Objetivos da aula

- Aprender sobre como importar e criar um Data Frame.



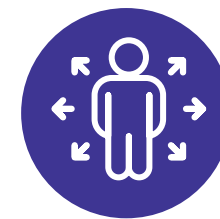
Recursos didáticos

- 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 com base 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 – Criando

Importe a biblioteca:

```
# Importando a biblioteca
```

```
import pandas as pd
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o **conceito**

Pandas – Criando

Em seguida, abra ou crie o Data Frame:

```
import pandas as pd

# Criando um DataFrame chamado primeiro_df com 2 colunas 'a' e 'b' e 2 valores 1 e 2
primeiro_df = pd.DataFrame({
    'a': [1, 2],
    'b': [1, 2]
})

# Exibindo o DataFrame
primeiro_df
```

	a	b
0	1	1
1	2	2

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Pandas – Criando

Veja outro exemplo de criação de Data Frame:

```
import pandas as pd

# Criando um dicionário com dados básicos
dados_basicos = {
    "ID": [1, 2],
    "Nome": ["Alice", "Bob"],
    "Idade": [28, 32],
    "Cidade": ["São Paulo", "Rio de Janeiro"]
}

# Criando o DataFrame usando os dados básicos
df_basicos = pd.DataFrame(dados_basicos)

# Exibindo o DataFrame
df_basicos
```

	ID	Nome	Idade	Cidade
0	1	Alice	28	São Paulo
1	2	Bob	32	Rio de Janeiro

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo
o **conceito**

Pandas – Criando

Veja outro exemplo de criação de Data Frame:

```
import pandas as pd

# Criando um dicionário com dados de saúde
dados = {
    "ID": [1, 2, 3, 4, 5],
    "Nome do Paciente": ["João Silva", "Maria Souza", "Ana Costa", "Lucas Martins", "Beatriz Santos"],
    "Idade": [30, 25, 40, 22, 35],
    "Sexo": ["M", "F", "F", "M", "F"],
    "Diagnóstico": ["Gripe", "Alergia", "Hipertensão", "Fratura no braço", "Ansiedade"],
    "Data da Consulta": ["2024-04-01", "2024-04-02", "2024-04-03", "2024-04-04", "2024-04-05"],
    "Tratamento Recomendado": ["Repouso e hidratação", "Antialérgicos", "Dieta e exercícios", "Imobilização", "Terapia"],
    "Custo do Tratamento (R$)": [150, 200, 300, 450, 500]
}

# Criando o DataFrame usando os dados
df_saude = pd.DataFrame(dados)

# Exibindo o DataFrame
df_saude
```

	ID	Nome do Paciente	Idade	Sexo	Diagnóstico	Data da Consulta	Tratamento Recomendado	Custo do Tratamento (R\$)
0	1	João Silva	30	M	Gripe	2024-04-01	Repouso e hidratação	150
1	2	Maria Souza	25	F	Alergia	2024-04-02	Antialérgicos	200
2	3	Ana Costa	40	F	Hipertensão	2024-04-03	Dieta e exercícios	300
3	4	Lucas Martins	22	M	Fratura no braço	2024-04-04	Imobilização	450
4	5	Beatriz Santos	35	F	Ansiedade	2024-04-05	Terapia	500

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Pandas – Criando

Veja outro exemplo de criação de Data Frame:

```
import pandas as pd

dados_futebol = {
    "ID": [1, 2, 3, 4, 5],
    "Nome do Jogador": ["Cristiano Ronaldo", "Lionel Messi", "Neymar Jr", "Kylian Mbappé", "Robert Lewandowski"],
    "Idade": [36, 34, 29, 22, 33],
    "Clube Atual": ["Manchester United", "Paris Saint-Germain", "Paris Saint-Germain", "Paris Saint-Germain", "Bayern Munich"],
    "Gols na Temporada": [24, 30, 19, 27, 41]
}

# Criando o DataFrame usando os dados de futebol
df_futebol = pd.DataFrame(dados_futebol)

# Exibindo o DataFrame
df_futebol
```

	ID	Nome do Jogador	Idade	Clube Atual	Gols na Temporada
0	1	Cristiano Ronaldo	36	Manchester United	24
1	2	Lionel Messi	34	Paris Saint-Germain	30
2	3	Neymar Jr	29	Paris Saint-Germain	19
3	4	Kylian Mbappé	22	Paris Saint-Germain	27
4	5	Robert Lewandowski	33	Bayern Munich	41

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o conceito

Pandas – Criando

Para criar Data Frames, usamos a função `pd.DataFrame()`.

`pd.DataFrame()`: esta é a função mais direta para criar um Data Frame a partir de um dicionário de listas, uma lista de dicionários ou até mesmo um NumPy array.

```
d = {'col1': [1, 2], 'col2': [3, 4]}  
df = pd.DataFrame(data=d)  
df
```

	col1	col2
0	1	3
1	2	4

```
d = {'col1': [0, 1, 2, 3], 'col2': pd.Series([2, 3], index=[2, 3])}  
pd.DataFrame(data=d, index=[0, 1, 2, 3])
```

	col1	col2
0	0	NaN
1	1	NaN
2	2	2.0
3	3	3.0

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



Colocando
em **prática**

Construindo um Data Frame

Com base no Data Frame fornecido abaixo, desenvolva o código no Jupyter Notebook, utilizando o Pandas, para gerar essa visão de tabela:

	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



Enviar no AVA

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.



© Getty Images

O que nós
**aprendemos
hoje?**

Então ficamos assim...

- 1** Aprendemos como criar um Data Frame;
- 2** Praticamos a criação de um Data Frame com a função `pd.DataFrame`.

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**

**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.



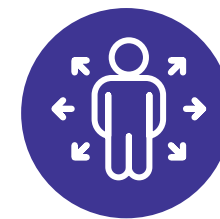
Recursos didáticos

- 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**