

**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.



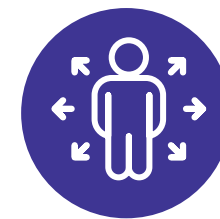
Recursos didáticos

- 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

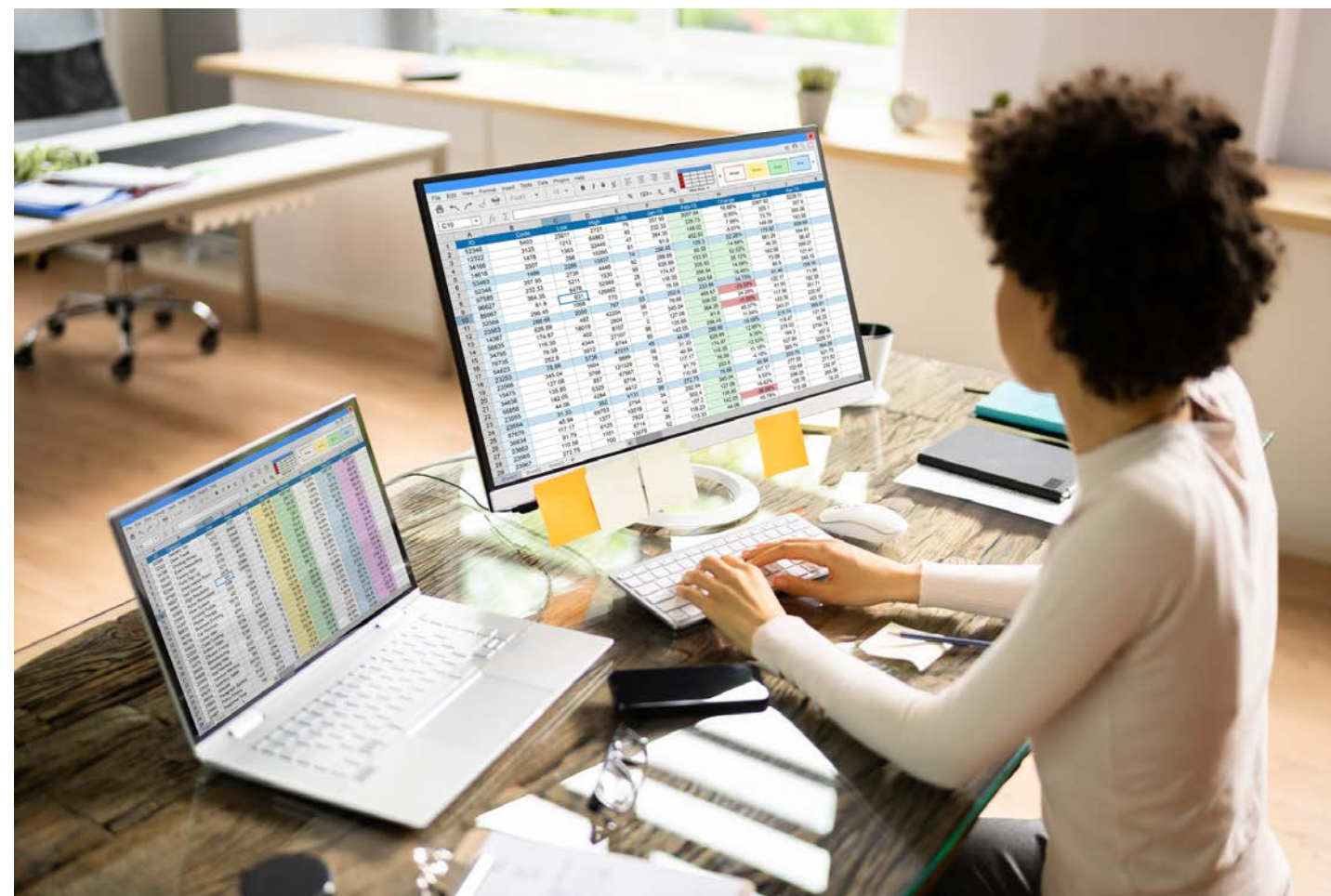
Elabore o relatório final

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.])

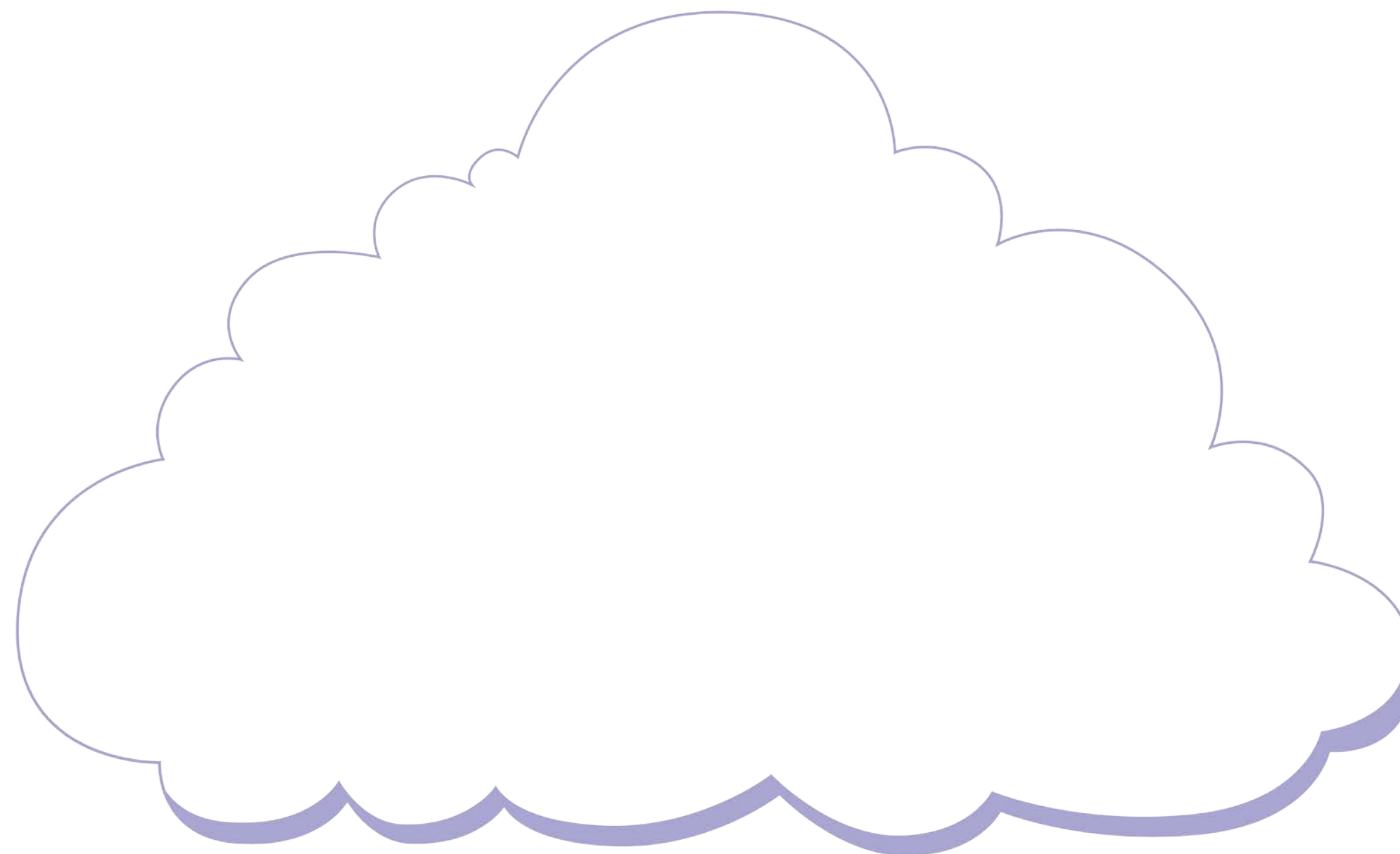
Construindo
o **conceito**

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**