

Educação Profissional Paulista

Técnico em
**Ciência de
Dados**

Introdução ao Jupyter Notebook e à configuração de ambiente

Introdução ao Jupyter Notebook

Trabalhando com dados e projetos no Jupyter Notebook

Código da aula: [DADOS]ANO1C1B4S28A3

Mapa da unidade 6 Componente 1

Introdução ao
Jupyter Notebook e à
configuração de
ambiente

semana

28

Você está aqui!
Introdução ao
Jupyter Notebook

semana

30

Principais
funcionalidades do
Jupyter Notebook

semana

31

Configuração de
ambiente para o
Jupyter Notebook

**Introdução ao
Jupyter Notebook e à
configuração de
ambiente**

Mapa da unidade 6 Componente 1

Você está aqui!

Introdução ao Jupyter Notebook

Aula 3

Código da aula: [DADOS]ANO1C1B4S28A3

28



Objetivos da aula

- Introduzir, configurar e utilizar o Jupyter Notebook.



Recursos didáticos

- Recurso audiovisual para exibição de vídeos e imagens;
- Acesso ao laboratório de informática e/ou à internet.



Duração da aula

50 minutos.



Competências técnicas

- Usar ferramentas de desenvolvimento de software como o Jupyter Notebook;
- Compreender e dominar técnicas de manipulação de dados.



Competências socioemocionais

- Analisar informações de forma crítica e tomar decisões;
- Avaliar a qualidade dos dados e a eficácia dos modelos e das técnicas de análise;
- 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**

Carregando e manipulando dados com Pandas

- ▶ **Pandas** é uma biblioteca importante do Python para a análise de dados. Ela oferece estruturas de dados como DataFrames, que são tabelas bidimensionais com rótulos nos eixos.
- ▶ Pandas facilita o carregamento, a manipulação e a análise de grandes conjuntos de dados de diversas fontes, incluindo:
 - ✓ CSV;
 - ✓ Excel;
 - ✓ SQL;
 - ✓ JSON.



Reprodução - PANDAS. [s.d.]. Disponível em:
<https://pandas.pydata.org>. Acesso em: 24 jul. 2024.

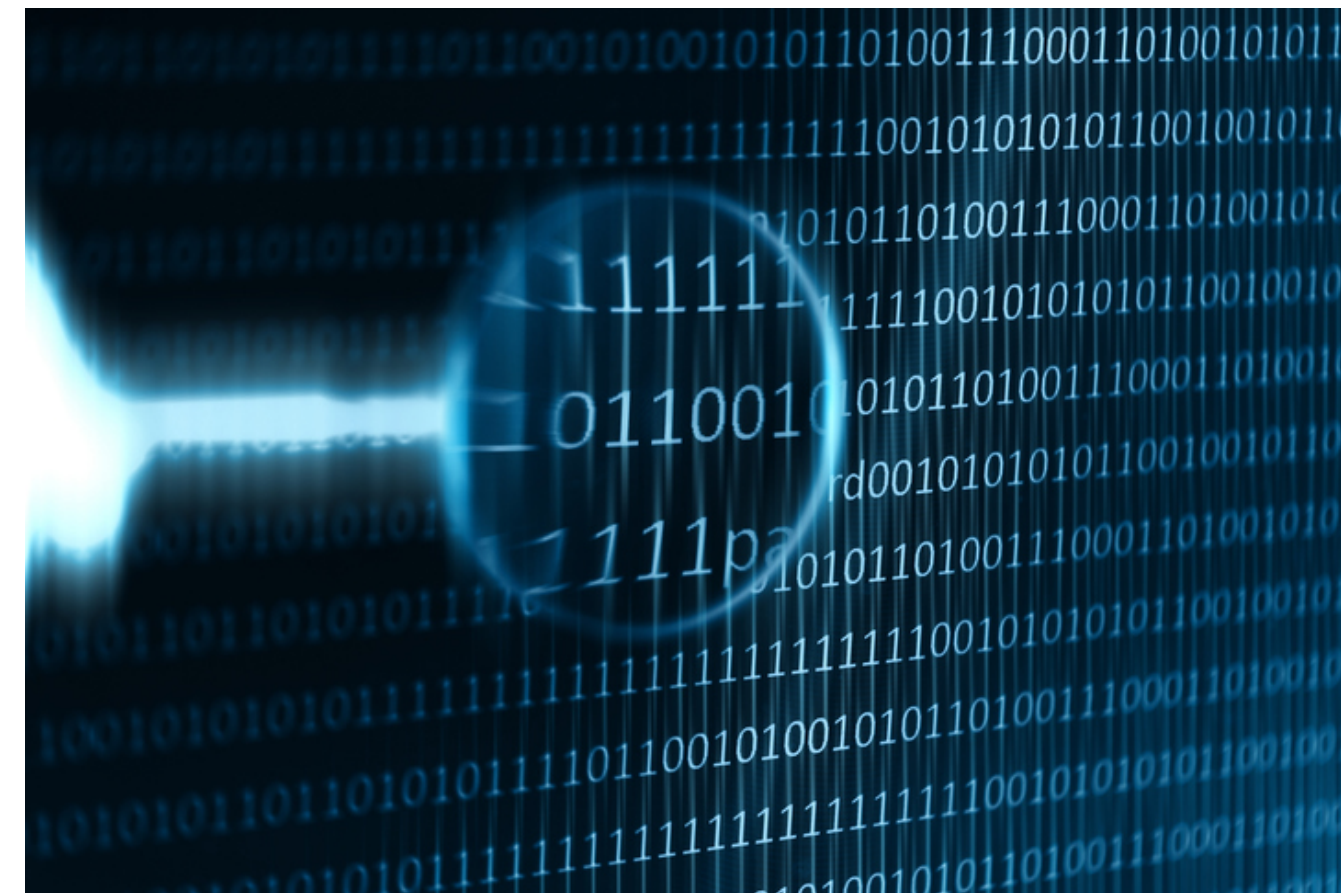
Construindo o conceito

Análise exploratória de dados (EDA)

A análise exploratória de dados (EDA) é uma abordagem para resumir e visualizar as principais características de um conjunto de dados.

Utilizando ferramentas como Pandas e bibliotecas de visualização, os analistas podem:

- ✓ identificar padrões;
- ✓ detectar anomalias;
- ✓ testar hipóteses;
- ✓ verificar suposições.



© Getty Images

Construindo
o **conceito**

Boas práticas e organização de projetos no Jupyter Notebook

Manter notebooks organizados e bem documentados é essencial para a colaboração e a manutenção de projetos. As boas práticas incluem:

- ✓ Adição de cabeçalhos claros;
- ✓ Uso de comentários;
- ✓ Uso de células de Markdown para explicações;
- ✓ Organização lógica do fluxo de trabalho;
- ✓ O versionamento de notebooks com Git pode ajudar a rastrear mudanças e colaborar de forma eficaz.

Construindo o **conceito**

Exemplos práticos

Criando o arquivo CSV e mostrando o DataFrame:

```
import pandas as pd

# Criando um DataFrame com dados de vendas
data = {
    'id_venda': [1, 2, 3, 4, 5],
    'categoria': ['Eletrônicos', 'Roupas', 'Eletrônicos', 'Alimentos', 'Roupas'],
    'unidades_vendidas': [1500, 300, 1200, 800, 400],
    'valor_venda': [30000, 4500, 24000, 16000, 6000]
}

df_vendas = pd.DataFrame(data)

# Salvando o DataFrame em um arquivo CSV
csv_path = 'vendas.csv'
df_vendas.to_csv(csv_path, index=False)

# Exibindo o DataFrame
df_vendas.head()
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo
o **conceito**

Exemplos práticos

Carregar e visualizar o Arquivo CSV:

```
import pandas as pd

csv_path = 'vendas.csv'

# Carregar o arquivo CSV
df_vendas_carregado = pd.read_csv(csv_path)

# Visualizar as primeiras linhas do DataFrame
df_vendas_carregado.head()
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo
o **conceito**

Exemplos práticos

Análise exploratória de dados (EDA)

Exemplo 1 – Visualizar a distribuição das vendas:

```
import seaborn as sns
import matplotlib.pyplot as plt

sns.histplot(df_vendas['valor_venda'])
plt.title('Distribuição das Vendas')
plt.show()
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo
o **conceito**

Exemplos práticos

Análise exploratória de dados (EDA)

Exemplo 2 – Analisar a relação entre unidades vendidas e valor de venda:

```
sns.scatterplot(x='unidades_vendidas', y='valor_venda', data=df_vendas)  
plt.title('Unidades Vendidas vs Valor de Venda')  
plt.show()
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo
o **conceito**

Exemplos práticos

Análise exploratória de dados (EDA)

Exemplo 3 – Identificar *outliers* em vendas.

```
sns.boxplot(x=df_vendas['valor_venda'])  
plt.title('Outliers em Valor de Venda')  
plt.show()
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Construindo o **conceito**

Exemplos práticos

Boas práticas e organização de projetos no Jupyter Notebook

Exemplo 1 – Adicionar cabeçalho e explicação usando Markdown:

```
# Análise de Vendas  
Este notebook contém a análise das vendas do último trimestre.
```

Exemplo 2 – Comentar o código para clarificar sua funcionalidade:

```
# Filtrar vendas acima de 1000 unidades  
vendas_altas = df_vendas[df_vendas['unidades_vendidas'] > 1000]
```

Elaborado especialmente para o curso com a ferramenta Jupyter Notebook.

Colocando
em **prática**

Exercício: utilizando o Jupyter



20 minutos



Em duplas

1

Crie um Jupyter Notebook que execute as seguintes tarefas:

2

Carregar dados: carregue o arquivo vendas.csv no Jupyter Notebook.

3

Visualizar dados: exiba as primeiras cinco linhas do DataFrame carregado;

Filtrar dados: filtre o DataFrame para mostrar apenas as vendas em que o valor de venda é maior que 10.000.

4

Agrupar dados: agrupe os dados por categoria e calcule a média de unidades vendidas para cada categoria.

5

Visualizar resultados: exiba o DataFrame resultante da operação de agrupamento.

Ser
sempre +

Situação

Maria é uma analista de dados recém-contratada em uma grande empresa de tecnologia. Ela foi designada para trabalhar em um projeto importante que envolve a análise de grandes volumes de dados de vendas.

Apesar de suas habilidades técnicas, Maria enfrenta dificuldades de comunicação com seu colega João, que é responsável pela parte de desenvolvimento do projeto.

João frequentemente entrega os dados atrasados, o que impede Maria de concluir suas análises a tempo. Maria está se sentindo frustrada e desmotivada, bem como está considerando conversar com seu gerente sobre a situação.

Situação fictícia elaborada especialmente para o curso.

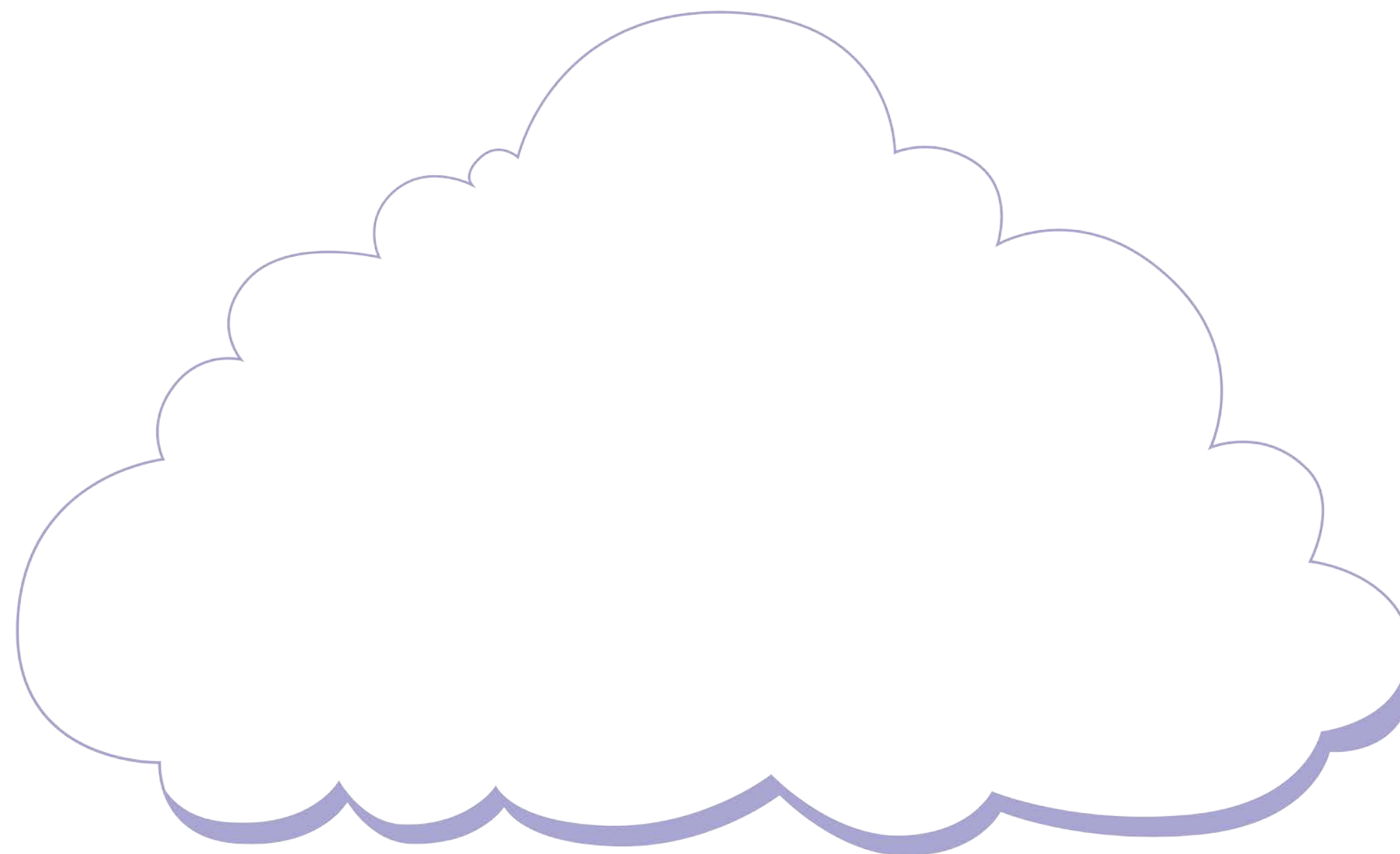
Ser
sempre +

Ação

Perguntas para reflexão e ação:

1. Qual é a importância de Maria conversar diretamente com João antes de levar a questão ao gerente?
2. Como Maria pode estruturar a conversa para que seja produtiva e profissional?
3. Quais habilidades socioemocionais Maria deve utilizar durante a conversa com João?

Nuvem de palavras



© Getty Images

O que nós
**aprendemos
hoje?**



© Getty Images

O que nós
aprendemos
hoje?

Então ficamos assim...

- 1** Aprendemos a utilizar a biblioteca Pandas para carregar e manipular dados. Exploramos operações fundamentais, como carregar arquivos CSV, filtrar dados e agrupar informações para análise.
- 2** Discutimos a análise exploratória de dados (EDA), utilizando Pandas e bibliotecas de visualização como Matplotlib e Seaborn. Visualizamos distribuições, relações entre variáveis e identificamos outliers.
- 3** Abordamos as boas práticas e a organização de projetos no Jupyter Notebook. Utilizamos células de Markdown para documentação, comentários no código para clareza e versionamento com Git para colaborar e rastrear mudanças.

Saiba mais

Jupyter Notebook é o queridinho da ciência de dados, mas já ouviu falar do JupyterLab? Essa versão mais moderna oferece recursos extras, como extensões!

MIRANDA, J. V. Conhecendo o JupyterLab. Alura, 16 ago. 2023. Disponível em:
<https://www.alura.com.br/artigos/conhecendo-jupyterlab>.
Acesso em: 23 jul. 2024.

Referências da aula

JUPYTER. *Página inicial*, [s.d.]. Disponível em: <https://jupyter.org>. Acesso em: 24 jul. 2024.

Identidade visual: Imagens © Getty Images

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**