

Semana 18 – Parte 1

Introdução à Ciência de Dados c/ Pandas e NumPy

 Manipulação de Dados com Python

 Curso: Ciência de Dados para Negócios – Fatec Votorantim

 Base: Capítulo 11 do livro Python Essencial

Objetivos da Aula

Fundamentos de Pandas e NumPy

Compreender o papel essencial dessas bibliotecas na análise de dados moderna.

Otimização com Arrays e DataFrames

Comparar o desempenho de listas e dicionários nativos com arrays e DataFrames.

Manipulação de Dados Essencial

Aprender a carregar, inspecionar e realizar operações básicas em DataFrames.

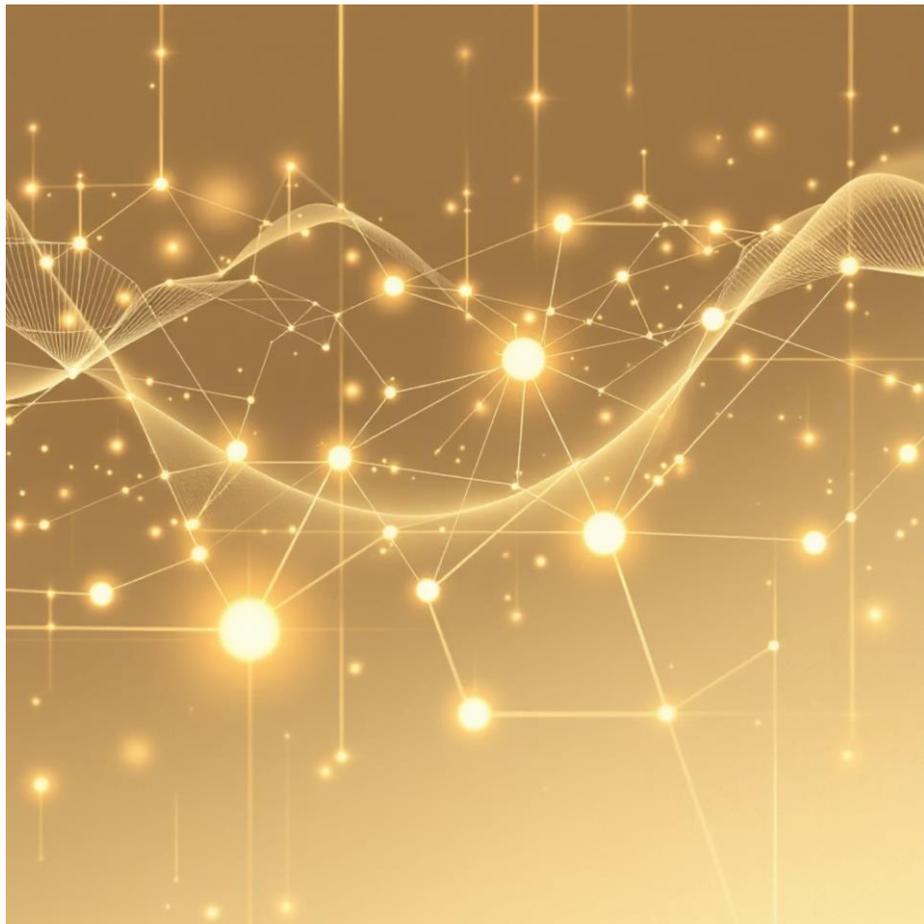
Pensamento Computacional Aplicado

Desenvolver habilidades de abstração, decomposição e análise para resolver problemas complexos.

🔧 O que são NumPy e Pandas?

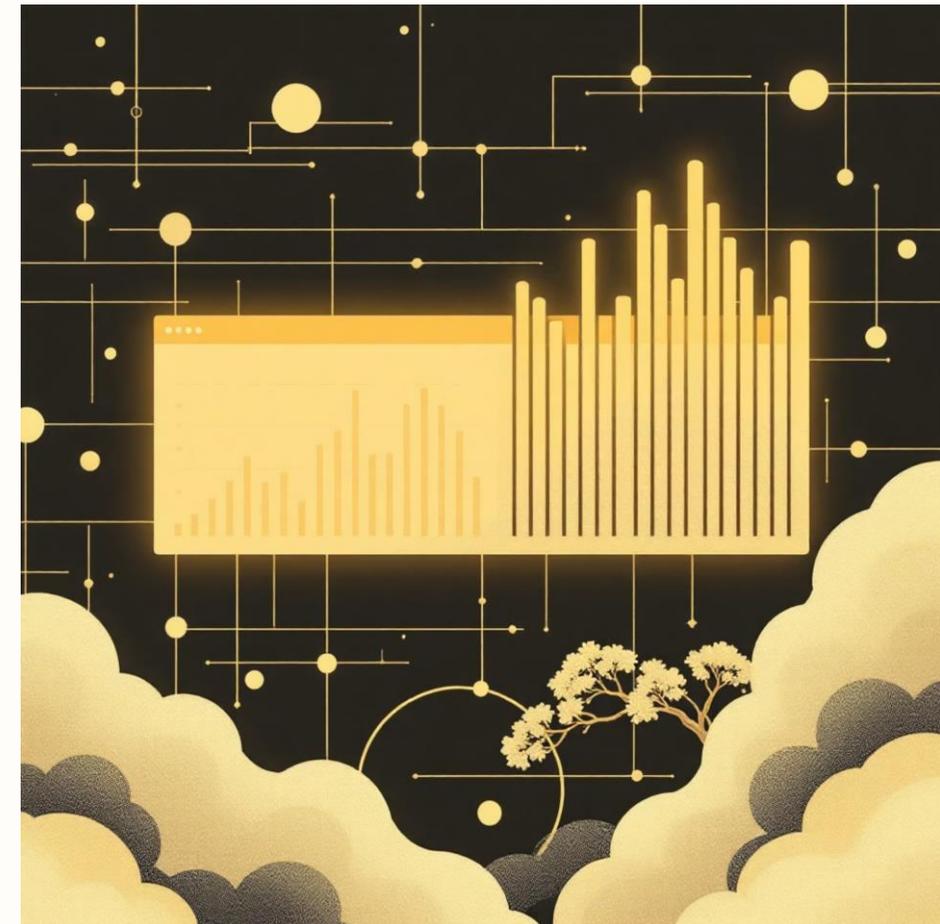
NumPy: O Poder Numérico

- Foco em performance e eficiência numérica.
- Trabalha com **arrays** multidimensionais para computação otimizada.
- Ideal para grandes volumes de cálculos numéricos e operações matemáticas complexas.



Pandas: A Análise Estruturada

- Foco em **análise de dados tabulares** e manipulação de conjuntos de dados.
- Oferece estruturas intuitivas como **Series** e **DataFrames**.
- Facilita tarefas como leitura de CSVs, agrupamentos, filtragens e transformações de dados.



DataFrame: A Tabela Inteligente do Pandas

Um **DataFrame** é a estrutura central do Pandas, funcionando como uma tabela flexível e poderosa para seus dados.

| | | | |
|------------|---------|-----|-------|
| 2025-08-01 | Caneta | 100 | 2.50 |
| 2025-08-01 | Caderno | 50 | 15.00 |
| 2025-08-02 | Mochila | 20 | 80.00 |

Acesso e Seleção

Acesse colunas facilmente:

```
df['Produto'].
```

Estatísticas Rápidas

Calcule estatísticas como a média:

```
df['Preço Unitário'].mean().
```

Criação de Novas Colunas

Adicione colunas derivadas:

```
df['Receita'] =  
df['Quantidade'] * df['Preço  
Unitário'].
```



Atividade Guiada: Mão na Massa!



Nesta atividade prática, utilizaremos o ambiente Jupyter/Colab para executar os seguintes passos:

- **Carregar Dados:** Abrir um arquivo CSV no DataFrame.
- **Inspecionar:** Visualizar as primeiras linhas com `df.head()`.
- **Filtrar:** Selecionar registros com base em condições, ex: `df[df['Valor'] > 100]`.
- **Criar Colunas:** Derivar novas informações a partir de dados existentes.
- **Agrupar:** Calcular médias e outras agregações com `groupby()`.
- **Tratar Ausentes:** Gerenciar valores NaN com `fillna()`.
- **Ordenar:** Reorganizar o DataFrame por datas com `df.sort_values('Data')`.

🔧 NumPy em Ação

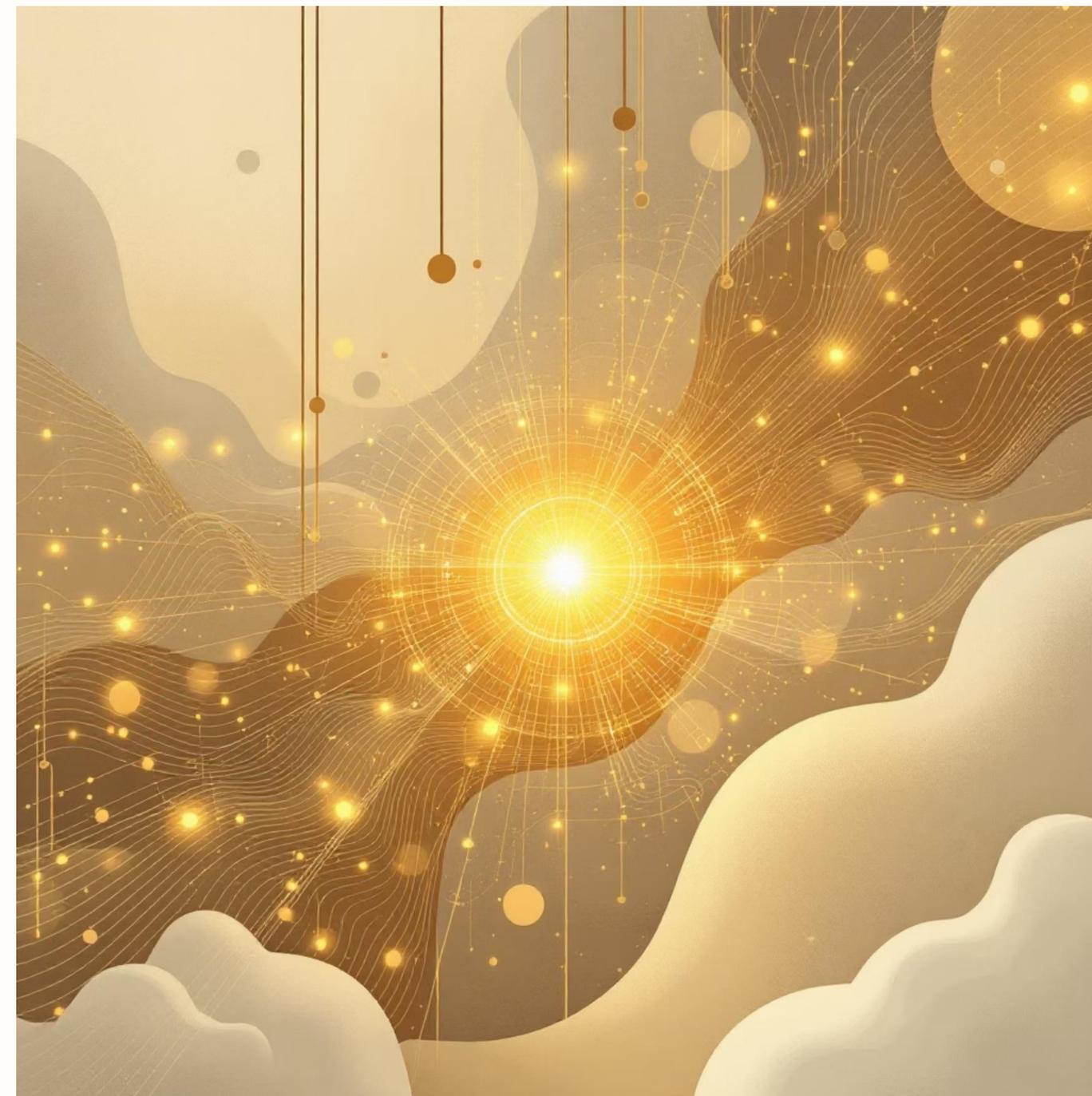
NumPy é a base para computação numérica em Python, oferecendo velocidade e eficiência incomparáveis.

```
import numpy as np
dados = np.array([1, 2, 3, 4, 5])
np.sum(dados) # Mais eficiente que laço for
# Resultado: 15
```

Por que usar NumPy?

- **Velocidade:** Operações com arrays são significativamente mais rápidas que com listas Python para cálculos numéricos.
- **Eficiência:** Menor consumo de memória para grandes conjuntos de dados.
- **Funcionalidades:** Oferece uma vasta gama de funções matemáticas otimizadas para arrays.

💡 Ideal para: Simulações complexas, álgebra linear, transformações de Fourier e manipulação vetorial em ciência de dados e machine learning.



Como a IA pode Ajudar?



Assistentes de Código

Ferramentas como **Copilot** e **Colab IA** sugerem comandos e completam seu código, acelerando o desenvolvimento.

Ex:

```
df.read_csv("arquivo.csv").
```



Explicação e Debug

Modelos como **ChatGPT**, **Gemini** e **DeepSeek** podem explicar erros, sugerir soluções para problemas complexos e até otimizar seu código. Ex: "Como agrupar por produto e calcular a média?".

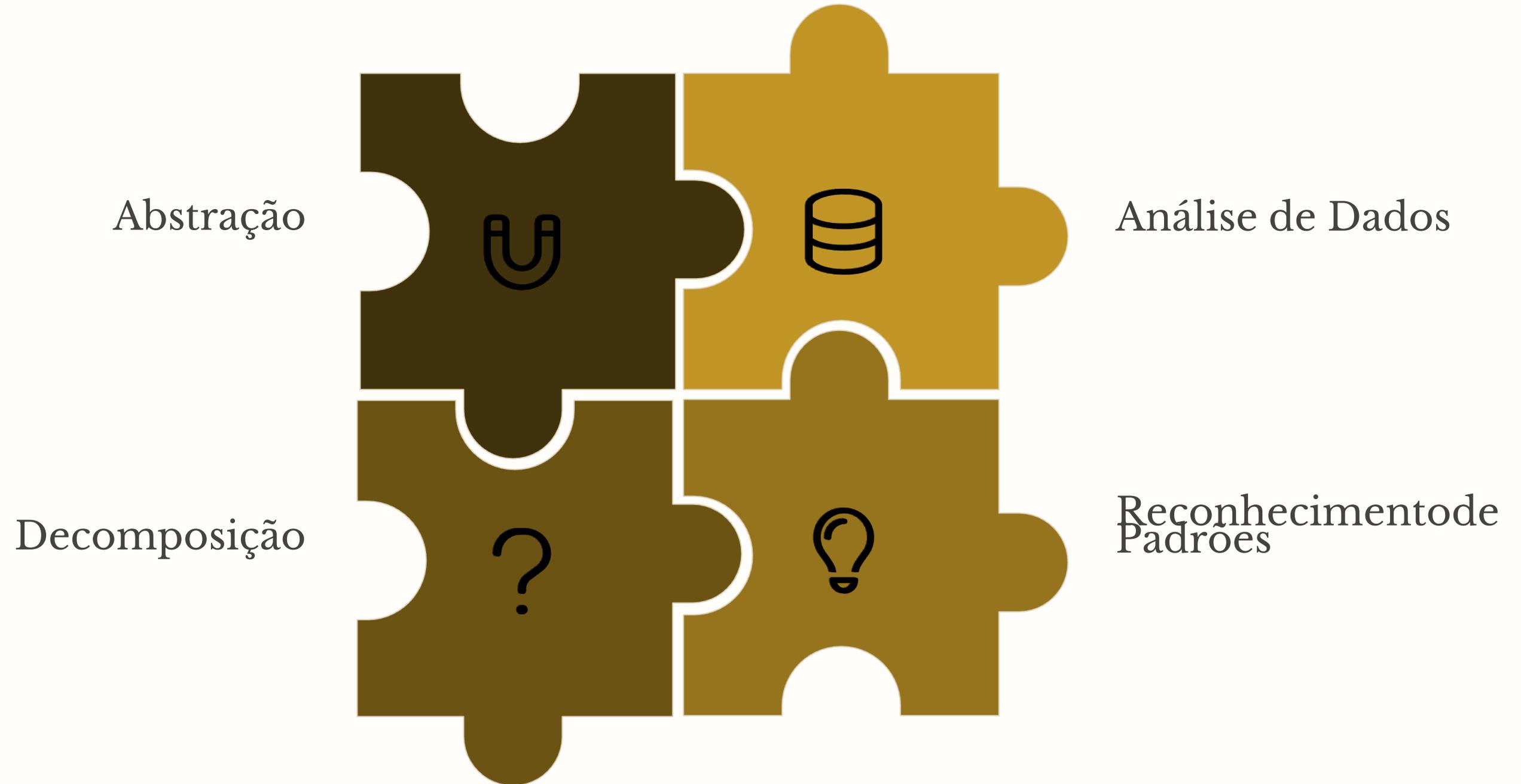


Tutor Auxiliar

A IA atua como um tutor que oferece insights e direções, mas o entendimento e a validação humana continuam sendo cruciais para a análise de dados eficaz.

Pensamento Computacional em Ação

O pensamento computacional é uma habilidade fundamental na ciência de dados, permitindo abordar problemas complexos de forma estruturada.



🧠 Taxonomia de Bloom: Aplicando e Analisando

Na Ciência de Dados, vamos além da memorização, focando na aplicação prática e na análise crítica dos resultados.

✅ Aplicar

Utilize os comandos de Pandas e NumPy para resolver tarefas reais de manipulação e análise de dados.



✅ Analisar

Interprete os resultados obtidos, questionando e buscando insights significativos.

- **Por que a média está distorcida?** (Detecte anomalias ou vieses)
- **Qual produto mais vendeu? O que isso nos diz?** (Tire conclusões de negócios)



Para Praticar e Explorar

A prática leva à maestria! O notebook da aula contém desafios essenciais para você consolidar seu aprendizado.



Análise de Vendas

Encontre o produto com a maior venda total.



Tratamento de Dados

Identifique e substitua dias com vendas zero por NaN.



Agrupamentos

Realize agrupamentos por produto para insights detalhados.



Ordenação

Ordene seus dados por data para visualizar tendências temporais.



Conversão de Tipos

Converta colunas de texto para o tipo DateTime.

 Faça os exercícios no seu ambiente Colab ou Jupyter. **Compartilhe suas descobertas e dúvidas!**

 **Estude antes:** Cap. 11 do livro + Módulo 11 do curso online pypro.com.br