1ª SEMANA

03-10-2025

Tópicos:

DataScience

Series e DataFrame

DataScience

**O que é Ciência de Dados?**

É a área que combina estatística, programação e conhecimento do negócio para extrair insights ou percepção (compreensão) de dados. Em outras palavras “Ciência de Dados é contar histórias usando dados, com ajuda de matemática e computadores.”

O objetivo é transformar dados brutos em decisões úteis.

Exemplo do dia a dia:

Netflix recomenda filmes analisando o histórico de visualização.

Bancos usam modelos para detectar fraude em tempo real.

Supermercados analisam vendas para prever demanda.

**Área de atuação**

Na prática, temos vários perfis dentro de Data Science:

Cientista de Dados (Data Scientist): cria modelos estatísticos e de machine learning para prever ou classificar.

Engenheiro de Dados (Data Engineer): organiza, processa e armazena os dados em grande escala.

Analista de Dados (Data Analyst): faz relatórios, dashboards, análises descritivas.

Machine Learning Engineer: coloca modelos em produção (apps, APIs).

Business Intelligence (BI): foca em relatórios para gestão.

**Mercado:**

Alta demanda no Brasil e no mundo.

Python e SQL são as skills mais cobradas em entrevistas.

Domínio de estatística básica e machine learning é um diferencial.

**Principais ferramentas:**

Python: linguagem principal para ciência de dados.

Jupyter Notebook: indicado para testes, ensino e experimentos.

VS Code: editor profissional, usado para projetos maiores.

Google Colab: ambiente online gratuito, ideal para sala de aula e começar sem instalar nada.

**Bibliotecas essenciais:**

NumPy: matemática com arrays.

Pandas: manipulação de tabelas (Excel no Python).

Matplotlib / Seaborn: gráficos e visualização.

Scikit-learn: machine learning.

Series

DataFrame

Primeiro exemplo:

1. Foi sugerido criar um mini conjunto de dados e sua transformação em um DataFrame

import pandas as pd

*#Criação de dicionário e lista para guardar desempenho dos alunos.*

*notaAlunos = {*

*"Nome": ["Pedro", "Igor", "Haylena", "Iasmin", "Helder"],*

*"NotaBigData": [9,8,8,9,7],*

*"NotaPython":[10,9,8.8, 9.5, 8.9]*

*}*

Nota. O conceito demonstrado cominação dicionário { } e lista [ ]

1..2 Para transformar o conjunto de dados em um DataFrame, basta aplicar a função pd.DataFrame e o nome da variável responsável por guardar dados. No nosso caso, foi notaAlunos

#Exemplo de criação

df = pd.DataFrame(notaAlunos)

#Exibir ou mostrar df convertido

print(df)

2. Próximo passo, se desejar pode transformar os nomes como índice, usando a funçao set\_index

#Exemplo

df\_com\_indece = df.set\_index(“Nome”)

3. É possível que queira já exportor o arquivo para ambiente externo em outros formatos, por exemplo .csv, .xlsx, entre outros. Para o efeito, aplica-se o método to\_csv(), to\_excel()

**OBS. CONTINUAÇÃO DISPONÍVEL NO ARQUIVO DA aula01, arquivo** **dataFrame.ipynb**