

Módulo 8: Backend com C# e Python

Dia 06: Python para Dados - ETL com Pandas

João Quintino

29 de dezembro de 2025



- 1 Ecossistema de Data Science
- 2 Pandas: DataFrames e Series
- 3 ETL: Leitura, Limpeza e Transformação
- 4 Prática: Processador de Relatórios
- 5 Encerramento

Onde entra o Python para Dados?

- Tratamento de grandes volumes de dados (CSV, Excel, SQL).
- Preparação de dados para Machine Learning.
- Geração de relatórios e dashboards.
- Automação de rotinas chatas de planilhas.

Stack de Dados em Python

- **Pandas:** DataFrames e manipulação tabular.
- **NumPy:** operações numéricas e vetoriais.
- **Matplotlib / Seaborn:** gráficos.
- **Scikit-learn:** Machine Learning clássico.

O que é um DataFrame?

Tabela em memória

```
import pandas as pd

dados = {
    "produto": ["Teclado", "Mouse"],
    "quantidade": [10, 5],
    "preco": [150.0, 80.0],
}

df = pd.DataFrame(dados)
print(df)
```

Lendo arquivos com Pandas

CSV e Excel

```
import pandas as pd

df_csv = pd.read_csv("vendas.csv")
df_excel = pd.read_excel("vendas.xlsx")

print(df_csv.head())
```

O que é ETL?

- **Extract:** ler dados de uma ou mais fontes.
- **Transform:** limpar, filtrar, calcular colunas.
- **Load:** salvar o resultado em outro formato (CSV, JSON, DB).

Removendo nulos e filtrando linhas

```
df = pd.read_excel("vendas.xlsx")  
  
df = df.dropna(subset=["produto", "preco"])  
df = df[df["quantidade"] > 0]
```


Criando colunas derivadas

Valor total da venda

```
df["valor_total"] = df["quantidade"] * df["preco"]

por_produto = df.groupby("produto")["valor_total"].sum()
print(por_produto)
```

Desafio Guiado: Excel → JSON

Objetivo

Ler um arquivo `vendas.xlsx` e gerar um `vendas_resumo.json` com:

- Total de vendas por produto.
- Quantidade total vendida por produto.

Fluxo geral do script

Função principal de ETL

```
import pandas as pd

def processar_vendas(caminho_excel: str, caminho_saida
: str) -> None:
    df = pd.read_excel(caminho_excel)
    # limpar
    # criar colunas
    # agrupar
    # salvar em JSON
```

- Ecossistema de Data Science com Python.
- DataFrames e Series com Pandas.
- Leitura de arquivos CSV/Excel.
- Limpeza e transformação básica de dados.
- Ideia de um pipeline ETL simples (Excel → JSON).

Vamos transformar dados!

- Adicionar mais campos ao relatório (data, categoria, etc.).
- Integrar essa etapa de ETL com as APIs futuras.