

Projeto de Análise de Dados com Python (Capítulos 1-7)

Baseado no livro *Python for Data Analysis* de Wes McKinney

Objetivo

Você trabalha como analista de dados em uma empresa de e-commerce. A empresa possui um conjunto de dados de vendas (`sales_data.csv`) e informações de clientes (`customers.csv`). Seu objetivo é explorar, limpar e analisar esses dados para extrair insights relevantes.

Conjunto de Dados

1. `sales_data.csv`

- `order_id`: ID único do pedido
- `customer_id`: ID do cliente
- `product_id`: ID do produto
- `order_date`: Data do pedido
- `quantity`: Quantidade comprada
- `unit_price`: Preço unitário do produto
- `total_price`: Preço total ($\text{quantity} \times \text{unit_price}$)

2. `customers.csv`

- `customer_id`: ID do cliente
- `name`: Nome do cliente
- `age`: Idade
- `gender`: Gênero (M/F)
- `city`: Cidade de residência

(Dados fictícios podem ser gerados aleatoriamente ou criados manualmente em CSV.)

Etapas do Projeto

1. Configuração Inicial (Cap. 2, 3)

- Importe as bibliotecas `pandas`, `numpy` e `matplotlib`.
- Carregue os arquivos `sales_data.csv` e `customers.csv` em `DataFrames`.
- Verifique a estrutura dos dados (`head()`, `info()`, `describe()`).

2. Limpeza e Pré-Processamento (Cap. 4, 5)

- Verifique e trate valores ausentes (`NaN`).

- Corrija inconsistências (ex.: datas mal formatadas, duplicatas).
- Converta tipos de dados se necessário (ex.: `order_date` para `datetime`).
- Adicione uma coluna `month` extraindo o mês de `order_date`.

3. Análise Exploratória (Cap. 6, 7)

- **Estatísticas Básicas:**
 - Número total de pedidos, clientes únicos e produtos.
 - Média, mediana e desvio padrão de `total_price`.
- **Agregações:**
 - Total de vendas por mês.
 - Top 5 clientes que mais gastaram.
 - Produtos mais vendidos (por quantidade).
- **Visualização:**
 - Gráfico de linha: Vendas mensais ao longo do tempo.
 - Gráfico de barras: Top 10 produtos por receita.
 - Histograma: Distribuição de idades dos clientes.

4. Merge e Análise Avançada (Cap. 5, 7)

- Combine `sales_data` e `customers` usando `customer_id`.
- Calcule métricas segmentadas por:
 - Gênero (`gender`): Qual gênero gasta mais em média?
 - Cidade (`city`): Quais cidades têm os clientes mais valiosos?
- Identifique padrões sazonais (ex.: vendas maiores em dezembro).

5. Relatório Final

- Crie um resumo executivo em um Jupyter Notebook com:
 - Principais descobertas.
 - Visualizações-chave.
 - Recomendações para a empresa (ex.: focar em produtos/cidades lucrativas).

Dicas

- Use `pd.merge()` para combinar DataFrames.
- Métodos úteis: `groupby()`, `agg()`, `value_counts()`, `pivot_table()`.
- Bibliotecas para visualização: `matplotlib.pyplot`, `seaborn`.

Bom projeto! 🚀 (Se precisar de ajuda pontual, posso dar orientações sem revelar soluções prontas.)