

Documentação do Projeto de BI: Data Warehouse com Python

Estrutura do Projeto

```
■■■ data/
■   ■■■ raw/          # CSVs brutos: produtos.csv, vendas.csv
■   ■■■ processed/    # CSV tratado: vendas_tratada.csv
■■■ utils/
■   ■■■ gera_produtos.py
■   ■■■ gera_vendas.py
■■■ etl/
■   ■■■ processa_vendas.py
■■■ dw/
■   ■■■ carrega_dw.py
■   ■■■ criar_dim_tempo.py
■   ■■■ criar_dim_produto.py
■   ■■■ criar_dim_loja.py
■■■ dw_vendas.db      # Banco de dados SQLite
■■■ main.py           # Script principal para executar toda a pipeline
```

Descrição das Etapas do Projeto

1. Geração de Dados Brutos

- Criação dos arquivos CSV com dados simulados de produtos e vendas.
- Scripts utilizados: gera_produtos.py, gera_vendas.py
- Esses arquivos geram os dados iniciais que serão utilizados na pipeline.

2. Processamento dos Dados (ETL)

- Tratamento e enriquecimento dos dados brutos.
- Exemplo: cálculo do total de vendas, limpeza e padronização.
- Resultado: arquivo CSV processado salvo em data/processed/vendas_tratada.csv.
- Script responsável: processa_vendas.py

3. Carga no Data Warehouse

- Importação dos dados tratados para o banco de dados SQLite.
- Script responsável: carrega_dw.py
- Banco de dados gerado: dw_vendas.db

4. Criação das Tabelas Dimensão

- Gerar tabelas dim_tempo, dim_produto e dim_loja a partir da tabela fato.
- Scripts responsáveis: criar_dim_tempo.py, criar_dim_produto.py, criar_dim_loja.py

5. Automatização da Pipeline

- Script principal main.py orquestra todas as etapas em sequência.

Fluxo do Script Principal (main.py)

1. Executa gera_produtos.py
2. Executa gera_vendas.py
3. Executa processa_vendas.py
4. Executa carrega_dw.py

5. Executa criar_dim_tempo.py
6. Executa criar_dim_produto.py
7. Executa criar_dim_loja.py