

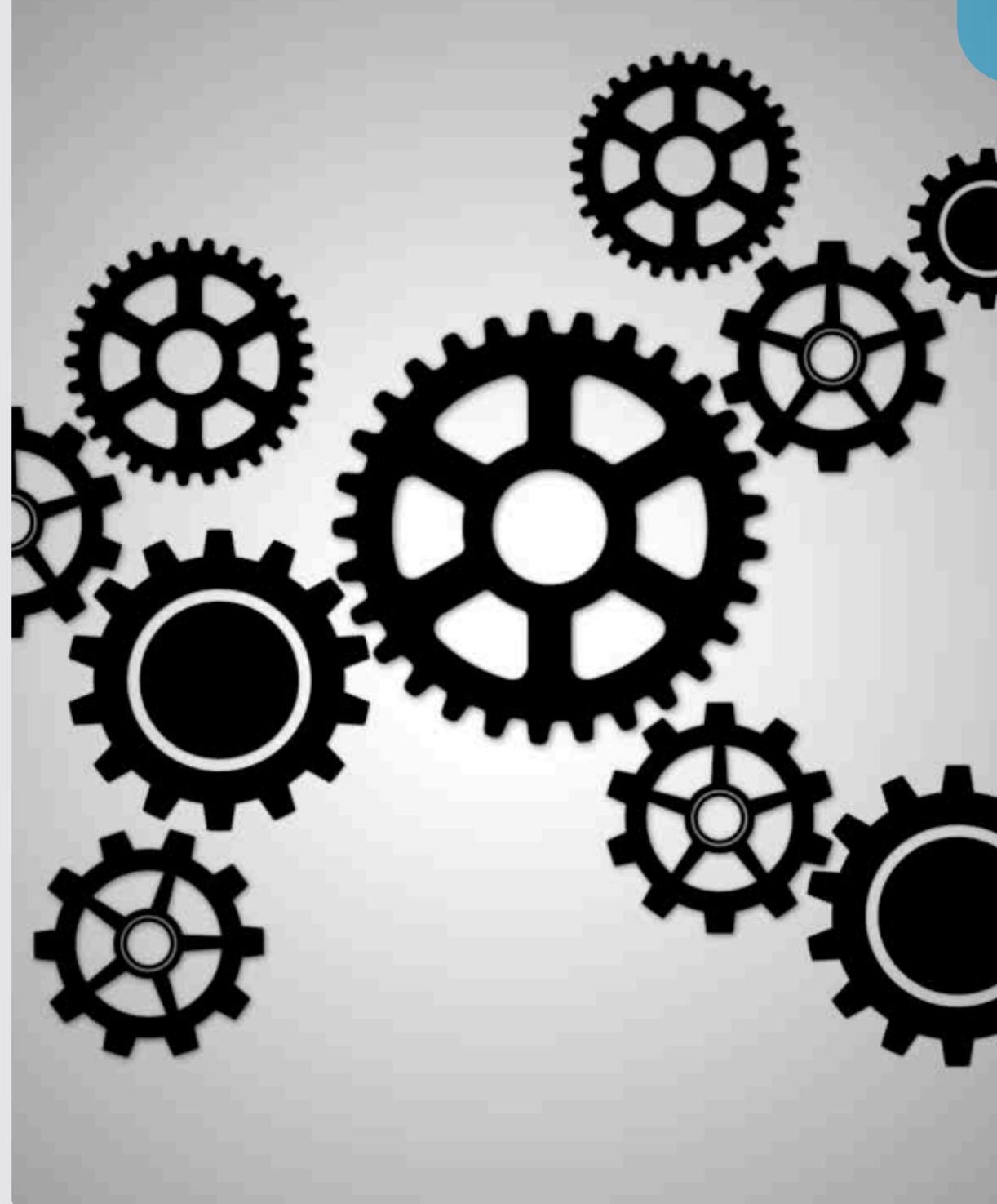
# PIPELINE DE DADOS

Prof. Wesley Andrade

21 de março de 2025



# AULA PRÁTICA VALENDO 20 PONTOS!



- APLICAÇÃO DA ETAPA DE EXTRAÇÃO DE DADOS
- APLICAÇÃO DAS TÉCNICAS DA ETAPA DE TRANSFORMAÇÃO DE DADOS
- APLICAÇÃO DA MODELAGEM DIMENSIONAL DE DADOS
- APLICAÇÃO DA ETAPA DE CARGA DE DADOS



# ATIVIDADE 01

## EXTRAÇÃO E EXPLORAÇÃO DOS DADOS



➤ Extraia os dados do arquivo 'dados\_vendas.csv'.  
Realize uma análise exploratória inicial mostrando as:

- dimensões (linhas e colunas);
- tipos de dados;
- valores ausentes; e
- resumo estatístico.



# ATIVIDADE 02

## IDENTIFICAÇÃO DE ENTIDADES



➤ Explore a estrutura dos dados identificando as entidades presentes, suas colunas correspondentes e a cardinalidade das relações.

- Identifique as entidades presentes no conjunto de dados.
- Para cada entidade, identifique as colunas correspondentes.
- Verifique a cardinalidade das relações analisando valores únicos nas colunas de ID.



# ATIVIDADE 03

## LIMPEZA DE DADOS

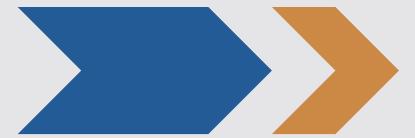


- Identifique e corrija os problemas nos dados: valores nulos, quantidades negativas, datas em formato errado, categorias inconsistentes e possíveis duplicatas.
- Trate os valores nulos nas colunas 'preco\_unitario', 'valor\_desconto' e 'nome\_fornecedor'.
  - Corrija as quantidades negativas.
  - Padronize todas as datas para o formato 'YYYY-MM-DD'.
  - Padronize as categorias de produtos (algumas estão em maiúsculas).
  - Identifique e trate possíveis duplicatas.



# ATIVIDADE 04

## CÁLCULO DE MÉTRICAS



➤ Calcule novas métricas para enriquecer a análise: valor bruto, valor líquido, margem de lucro e dias desde o pedido.

- Valor bruto (quantidade \* preco\_unitario)
- Valor líquido (valor bruto - valor\_desconto)
- Margem de lucro ((preco\_unitario - preco\_custo) / preco\_unitario \* 100)
- Dias desde o pedido (considerando a data atual)



# ATIVIDADE 05

## PROJETO DE MODELO DIMENSIONAL



- Identifique e corrija os problemas nos dados: valores nulos, quantidades negativas, datas em formato errado, categorias inconsistentes e possíveis duplicatas.
- Desenhe um diagrama conceitual identificando a tabela fato e as dimensões.
  - Defina as chaves primárias e estrangeiras.
  - Determine o nível de granularidade da tabela fato.
  - Identifique métricas (medidas) e atributos (dimensões).



# ATIVIDADE 06

## CRIAÇÃO DA DIMENSÃO DE PRODUTOS



- Implemente a dimensão de produtos (Dim\_Produto) selecionando apenas as colunas relacionadas a produtos, removendo duplicatas e organizando hierarquias.
- Selecione apenas as colunas relacionadas a produtos do dataset.
  - Remova duplicatas para garantir que cada produto apareça apenas uma vez.
  - Organize hierarquias (categoria > subcategoria).
  - Defina a chave primária.



# ATIVIDADE 07

## CRIAÇÃO DA DIMENSÃO DE CLIENTES



➤ Implemente a dimensão de clientes (Dim\_Cliente) selecionando as colunas relacionadas a clientes, removendo duplicatas e organizando hierarquias de localização.

- Selecione apenas as colunas relacionadas a clientes.
- Remova duplicatas.
- Organize hierarquias de localização.
- Defina a chave primária.



# ATIVIDADE 08

## CRIAÇÃO DA DIMENSÃO DE LOJAS



➤ Implemente a dimensão de lojas (Dim\_Loja) selecionando as colunas relacionadas a lojas, removendo duplicatas e organizando hierarquias geográficas.

- Selecione apenas as colunas relacionadas a lojas.
- Remova duplicatas.
- Organize hierarquias geográficas (região > estado > cidade).
- Defina a chave primária.



# ATIVIDADE 09

## CRIAÇÃO DA DIMENSÃO DE TEMPO



- Crie uma dimensão de tempo (Dim\_Tempo) a partir das datas de pedido extraindo componentes de data e adicionando indicadores úteis para análises.
- Extraia componentes de data: ano, trimestre, mês, semana, dia, dia da semana.
  - Adicione indicadores para fins de semana, feriados (simule alguns).
  - Crie classificações por período (manhã, tarde, noite) para análises futuras.
  - Considere a criação de campos para análises de sazonalidade.



# ATIVIDADE 10

## CRIAÇÃO DA TABELA FATO (VENDAS)



- Implemente a tabela fato de vendas (Fato\_Vendas) selecionando as métricas numéricas, incluindo as chaves estrangeiras para todas as dimensões e mantendo apenas os dados necessários.
- Selecione as métricas numéricas (quantidade, valores, etc.).
  - Inclua as chaves estrangeiras para todas as dimensões.
  - Mantenha apenas os dados necessários no nível de granularidade definido.
  - Valide a integridade referencial com as dimensões criadas.



# ATIVIDADE 11

## CRIAÇÃO DO BANCO DE DADOS (DW)



➤ Construa o esquema no SQLite para o data warehouse, definindo tabelas, chaves primárias, chaves estrangeiras e índices para melhorar performance.

- Crie o banco de dados 'dw\_vendas.sqlite'.
- Defina o esquema para todas as tabelas dimensionais e de fatos.
- Estabeleça chaves primárias e estrangeiras.



# ATIVIDADE 12

## CARGA NO DATA WAREHOUSE



➤ Carregue os dados processados no banco de dados SQLite, primeiro as dimensões e depois as tabelas fato, validando a integridade dos dados carregados.

- Carregue todas as dimensões primeiro.
- Carregue a tabela fato principal.



# PIPELINE DE DADOS

Prof. Wesley Andrade