

# LIMPEZA, INTEGRAÇÃO E TRANSFORMAÇÃO DE DADOS

Pré-processamento de Dados

25 Sep, 2025

## DESAFIO: LIMPEZA, INTEGRAÇÃO E TRANSFORMAÇÃO DE DADOS

### Objetivo

Esta tarefa tem como objetivo praticar as técnicas de preparação de dados para análise. Você recebeu 4 conjuntos de dados em diferentes formatos (CSV(2), Excel, JSON) que precisam ser integrados e preparados para análise. Conjuntos de Dados

1. **clientes.csv** - Informações dos clientes
2. **vendas.csv** - Registros de vendas
3. **produtos.xlsx** - Catálogo de produtos
4. **avaliacoes.json** - Avaliações dos produtos pelos clientes

### Desafios a Serem Identificados e Resolvidos

#### Problemas de Qualidade dos Dados

- Valores missing/ausentes
- Valores inconsistentes (ex: diferentes representações para mesma informação)
- Valores fora do domínio esperado
- Dados duplicados
- Formatos de data inconsistentes
- Problemas de encoding/caracteres especiais

#### Problemas de Integração

- Chaves primárias e estrangeiras inconsistentes
- Diferentes granularidades dos dados
- Conflitos de esquema entre datasets
- Dados semanticamente equivalentes com representações diferentes

#### Problemas de Transformação

- Necessidade de agregações
- Criação de novas features/colunas
- Normalização/padronização de dados
- Tratamento de outliers

## **Tarefas a serem realizadas**

### **Exploração e Diagnóstico dos dados**

1. Carregue cada conjunto de dados
2. Realize uma análise exploratória inicial
3. Identifique e documente todos os problemas encontrados

### **Limpeza e Transformação dos dados**

1. Corrija os problemas identificados em cada dataset individualmente
2. Padronize formatos e convenções
3. Trate valores missing, inconsistentes e outliers

### **Integração**

1. Una os datasets em uma base consolidada
2. Resolva conflitos de chaves e relacionamentos
3. Verifique a consistência dos dados integrados

### **Exploração mostre uma visualização dos dados (Graficos Varios dos dados integrados)**

### **Critérios de Avaliação da tarefa**

- Compreensão dos problemas de qualidade dos dados
- Aplicação adequada das técnicas de limpeza
- Eficácia na integração dos datasets
- Documentação do processo e decisões tomadas

### **Dicas**

- Documente cada etapa do processo
- Justifique suas decisões de tratamento
- Mantenha os dados originais preservados
- Valide os resultados após cada transformação

Bom trabalho!