



Contexto:

Uma empresa utiliza um sistema CRM (*Customer Relationship Management*) para gerir os seus clientes e vendas. Após vários anos de utilização, os dados ficaram inconsistentes e com erros. Como analista de dados, a tua tarefa é avaliar e melhorar a qualidade dos dados antes de integrares as tabelas num *Data Warehouse*.

Objetivos:

- Identificar e corrigir problemas de qualidade de dados (valores nulos, duplicados, tipos incorretos, idades inválidas).
- Validar regras de integridade referencial entre tabelas.
- Aplicar técnicas de limpeza e normalização.
- Gerar ficheiros limpos para análise posterior.

Ficheiros de trabalho:

- clientes_T5.csv
- vendas_T5.csv

Tarefas:

1. Leitura dos dados

- Lê os ficheiros clientes_T5.csv e vendas_T5.csv.

2. Detecção de problemas

- Verifica valores nulos, duplicados e idades fora do intervalo 18-100.
- Procura emails mal formatados (sem @ ou domínio).
- Verifica se todos os id_cliente das vendas existem na tabela de clientes.

3. Tratamento de dados

- Remove duplicados e registos inválidos.
- Corrige emails e preenche valores nulos.
- Elimina vendas sem cliente associado.

4. Normalização

- Formata nomes (str.title()).
- Coloca emails em minúsculas.

- Ordena clientes por id_cliente.

5. Exportação

- Gera os ficheiros clientes_final.csv e vendas_validadas.csv.

Nota (extra não obrigatória):

Implementar validação automática usando expressões regulares para:

- Emails
- Nomes (apenas letras e espaços)