



### Exercício 1

Temos a seguinte lista de dados de idades: [28, 45, -5, None, 32, 122, 98]. Quais valores devem ser marcados como inválidos?

### Exercício 2

Numa base de dados, dois clientes possuem o mesmo NIF mas nomes ligeiramente diferentes: “Marta Silva” e “Marta A. Silva”. O que investigar?

### Exercício 3

Considere a seguinte tabela de dados de clientes:

ClienteID	Nome	Email	Data Nascimento	Idade	Cidade	Saldo Conta
101	Ana Silva	<a href="mailto:ana.silva@email.com">ana.silva@email.com</a>	1990-05-15	33	Lisboa	1500,5€
102	Pedro Santos	<a href="mailto:pedro.santos@email.com">pedro.santos@email.com</a>	1985-11-20	38	Porto	2300€
103	Marta Costa		2000-01-01	150	Faro	500,25€
104	João Pereira	<a href="mailto:joao.pereira@email.com">joao.pereira@email.com</a>	2026-03-10	25	Lisboa	1800,75€
105	Ana Silva	<a href="mailto:ana.silva@email.com">ana.silva@email.com</a>	1990-05-15	33	Lisboa	1500,50€
106	Luísa F	<a href="mailto:carlos.gomes@email.com">carlos.gomes@email.com</a>	1995-07-22		Coimbra	900€
107	Carlos Gomes	maria@email.com	1978-02-28	45	PORTO	3000€
108	Maria		1992-10-01	31	Lisboa	-200€

Identifique pelo menos 5 problemas de qualidade de dados presentes nesta tabela, indicando a dimensão da qualidade afetada.

#### Exercício 4

Escreva as instruções SQL para criar uma tabela Produtos com as seguintes regras de validação:

- ProdutoID: Chave primária, não pode ser nulo e deve ser único.
- NomeProduto: Não pode ser nulo.
- Preço: Deve ser um valor numérico positivo (maior que 0).
- Stock: Deve ser um valor numérico não negativo (maior ou igual a 0).
- DataValidade: Se existir, deve ser uma data no futuro.

#### Exercício 5

Considere o ficheiro vendas.csv com os seguintes dados:

```
ID_Venda,Produto,Quantidade,Preco_Unitario,Data_Venda,Cliente_Email
1,Teclado,2,25.00,2024-01-10,cliente1@email.com
2,Rato,1,15.50,2024-01-11,cliente2@email.com
3,Monitor,1,150.00,2024-01-12,
4,Teclado,3,25.00,2024-01-10,cliente1@email.com
5,Webcam,1,30.00,2024-01-13,cliente3@email.com
6,Monitor,1,150.00,2024-01-12,cliente4@email.com
7,Headphones,0,50.00,2024-01-14,cliente5@email.com
8,Teclado,2,25.00,2024-01-10,cliente1@email.com
```

Escreva um script Python usando Pandas para realizar as seguintes operações de limpeza de dados:

- Carregar o ficheiro vendas.csv.
- Remover linhas duplicadas.
- Preencher valores em falta na coluna Cliente\_Email com "desconhecido@dominio.com".
- Corrigir valores inválidos na coluna Quantidade (deve ser maior que 0). Se Quantidade for 0 ou negativo, substituir por 1.
- Padronizar a coluna Produto para que a primeira letra de cada palavra seja maiúscula e as restantes minúsculas (e.g., "teclado" -> "Teclado").