**CASE TÉCNICO: ANALISTA DE SISTEMAS – LIMPEZA DE BASE DE DADOS**

**Contexto**

Você é um analista de sistemas em uma empresa que processa bases de dados em formatos CSV e Excel. Essas bases contêm caracteres especiais (como @, #, $, %, etc.) em campos de texto, o que causa problemas em sistemas internos. Sua tarefa é desenvolver um script Python que limpe esses caracteres especiais, mantendo apenas letras (a-z, A-Z), números (0-9), sublinhados (\_) e espaços.

**Objetivo**

Desenvolver um script Python que:

• Leia um arquivo de entrada no formato CSV ou Excel (.xlsx).

• Remova caracteres especiais de todas as colunas de texto, preservando letras, números, sublinhados e espaços.

• Remova espaços extras em campos de texto (ex.: “ texto com espaços ” → “texto com espaços”).

• Salve o resultado em um novo arquivo no mesmo formato do arquivo de entrada.

• Trate erros, como arquivos não encontrados ou formatos inválidos.

**Requisitos**

1. O script deve ser executado via linha de comando, recebendo dois argumentos:

• Caminho do arquivo de entrada (CSV ou Excel).

• Caminho do arquivo de saída.

2. Use a biblioteca pandas para manipulação de dados.

3. A limpeza deve ser aplicada apenas a colunas de texto.

4. O script deve ser robusto, com tratamento de erros.

5. Use nomes de variáveis em português e siga boas práticas de programação.

**Arquivos de Entrada**

Você receberá dois arquivos de exemplo:

• clientes.csv: Dados de clientes com colunas nome, idade, endereco e observacao.

• produtos.xlsx: Dados de produtos com colunas produto, preco, descricao e categoria.

Os arquivos contêm caracteres especiais em campos de texto. O script deve processar ambos os formatos.

**Entrega**

• Um script Python chamado limpeza\_dados.py.

• Os arquivos de saída gerados: clientes\_limpos.csv e produtos\_limpos.xlsx.

• Um breve comentário no início do script explicando sua abordagem.

**Exemplo de Uso**

python limpeza\_dados.py clientes.csv clientes\_limpos.csv

python limpeza\_dados.py produtos.xlsx produtos\_limpos.xlsx

**Critérios de Avaliação**

• Funcionalidade: O script remove caracteres especiais e espaços extras corretamente? Funciona para CSV e Excel?

• Robustez: O script lida com erros de forma adequada?

• Boas Práticas: O código é claro, com nomes de variáveis significativos e estrutura organizada?

• Eficiência: O script é otimizado para processar os dados?