



# TESTE DE SELEÇÃO

- DATA ENGINEER // OTO CRM -



## **// SOBRE A PMWEB**

A Pmweb é líder em Marketing Cloud Services e referência em Data Driven Marketing. Levamos nossos clientes a um novo patamar de resultados através de relacionamentos únicos em escala. Somos uma empresa de serviços e tecnologia para marketing e CRM.

Trabalhamos com clientes de cenários ultra-competitivos, dentre os principais players do varejo, turismo, educação e telecomunicação do país. Há mais de 25 anos no mercado, somos parceiros da Oracle, Sales Force, Google e Adobe. Fazemos parte do grupo WPP, a maior multinacional de publicidade e relações públicas do mundo.



## // SOBRE O TESTE

### O QUE DEVE SER FEITO

Você deverá implementar o fluxo de tratamento de dados descrito abaixo, utilizando o dataset a seguir como base:

<https://oto-public.s3.amazonaws.com/natal2021.zip>

#### **Sua tarefa:**

Com o objetivo de realizar uma campanha especial de final de ano, recebemos o arquivo `natal2021.csv` do cliente. Você recebeu o desafio de realizar a limpeza deste arquivo CSV para que possamos posteriormente importá-lo em nosso banco de dados. Pensando em um cenário no qual o arquivo CSV seja muito maior e não entre na memória de uma só vez, temos os seguintes problemas para serem resolvidos:

1. Identificar o encoding do arquivo;
2. Processar o arquivo de 1000 em 1000 linhas;
3. Remover os espaços das colunas. Ex. ' Porto Alegre ' -> 'Porto Alegre';
4. Criar uma coluna `CITY_ASCII` no arquivo, a qual deve ser construída com base na coluna `CITY`. Esta coluna não pode conter acentos, minúsculas e caracteres especiais. Apenas letras, números e hífen são permitidos. Ex. 'São Paulo - abc' -> 'SAO PAULO - ABC';
5. Remover os caracteres não numéricos da coluna `PHONE`;
6. Salvar o arquivo `.csv` com a nova coluna em UTF-8;



## O QUE SERÁ AVALIADO?

- Funcionalidade: se atinge o objetivo esperado dentro das limitações impostas.
- Legibilidade do código: será verificado se está fácil de ser lido e entendido, com nomes das funções e variáveis de acordo com a sua funcionalidade.
- Documentação, comentários e reaproveitamento de código.
- Organização na estrutura de arquivos.
- Boas práticas de desenvolvimento.
- Tempo de entrega.

## ENTREGAS NECESSÁRIAS

- Você deve salvar o código fonte necessário em um repositório Git da sua escolha.
- Enviar para o e-mail [pedro@pmweb.com.br](mailto:pedro@pmweb.com.br).