



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

O teste está dividido em duas etapas. Ambas as etapas foram pensadas tanto para que possamos conhecer um pouco mais das suas experiências técnicas e quanto para que você possa entender um pouco mais sobre os desafios que encontramos aqui na área de Dados.

ETAPA I – Python

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 - abç' -> 'SAO PAULO - ABC'
5. Remover os caracteres não numéricos da coluna `PHONE`;



6. Salvar o arquivo .csv com a nova coluna em UTF-8;

ETAPA II – SQL

Sua tarefa:

O Oto disponibiliza diversas métricas de resultados as quais são utilizadas pelos nossos clientes e recebem diariamente diversas requisições. Em um incidente recente, notou-se que a métrica de **quantidade de contatos realizados por região** estava performando de forma insatisfatória.

```
SELECT DISTINCT area, count()
FROM (SELECT pid, area FROM export AS c
INNER JOIN (
SELECT customer_fk
FROM contact
GROUP BY customer_fk
HAVING count() >= 1
) AS f ON c.pid = f.customer_fk
LEFT JOIN (SELECT DISTINCT store_state,
CASE
    WHEN store_state = 'RR' THEN 'Norte'
    WHEN store_state = 'AP' THEN 'Norte'
    WHEN store_state = 'PB' THEN 'Nordeste'
    WHEN store_state = 'AM' THEN 'Norte'
    WHEN store_state = 'PE' THEN 'Nordeste'
    WHEN store_state = 'AC' THEN 'Norte'
    WHEN store_state = 'MT' THEN 'Centro-oeste'
    WHEN store_state = 'RO' THEN 'Norte'
    WHEN store_state = 'MA' THEN 'Nordeste'
    WHEN store_state = 'PA' THEN 'Norte'
    WHEN store_state = 'PI' THEN 'Nordeste'
    WHEN store_state = 'GO' THEN 'Centro-oeste'
```



```
WHEN store_state = 'CE' THEN 'Nordeste'
WHEN store_state = 'DF' THEN 'Centro-oeste'
WHEN store_state = 'RN' THEN 'Nordeste'
WHEN store_state = 'AL' THEN 'Nordeste'
WHEN store_state = 'TO' THEN 'Norte'
WHEN store_state = 'SE' THEN 'Nordeste'
WHEN store_state = 'BA' THEN 'Nordeste'
WHEN store_state = 'MS' THEN 'Centro-oeste'
WHEN store_state = 'MG' THEN 'Sudeste'
WHEN store_state = 'ES' THEN 'Sudeste'
WHEN store_state = 'SC' THEN 'Sul'
WHEN store_state = 'RJ' THEN 'Sudeste'
WHEN store_state = 'SP' THEN 'Sudeste'
WHEN store_state = 'PR' THEN 'Sul'
WHEN store_state = 'RS' THEN 'Sul'
END AS area
FROM store_detail
ORDER BY area
) AS a
ON c.state = a.store_state
) GROUP BY area
```

Dada a estrutura de tabelas

```
export (
  pid BIGINT
  state STRING
  city STRING
  name STRING
  email STRING
  pref_store STRING
)
```



```
store_detail (  
    store_id BIGINT  
    store_name STRING  
    store_city STRING  
    store_state STRING  
)
```

```
contact (  
    customer_fk BIGINT  
    create_date DATE  
)
```

1. Faça as alterações necessárias na query para melhorar tanto a performance quanto a leitura da query e documente a razão de cada alteração.
2. Sugira um caminho para medir a melhoria de performance entre a query inicial e a versão desenvolvida.



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.