

Desafio Técnico para Data Engineer

Objetivo

O objetivo deste desafio é avaliar sua capacidade de, conscientemente, avaliar, propor e aplicar boas práticas e relacionadas à engenharia de dados para resolver um problema representativo. Especificamente, você deve processar dados do departamento de meteorologia holandês para identificar ondas de calor (e ondas de frio) na Holanda com base na definição local.

Contexto

Na Holanda, há uma definição bem clara acerca da ocorrência de uma onda de calor e de uma onda de frio, conforme explicado a seguir:

- **Onda de calor:** tenha-se registrado na localidade de **De Bilt** pelo menos 5 dias consecutivos com temperatura máxima diária maior ou igual a 25 °C, sendo que ao menos 3 destes dias (não necessariamente consecutivos), a temperatura máxima diária seja maior ou igual a 30 °C.
- **Onda de frio:** tenha-se registrado na localidade de **De Bilt** pelo menos 5 dias consecutivos com temperatura mínima diária menor ou igual a 0 °C sendo que ao menos 3 destes dias (não necessariamente consecutivos), a temperatura mínima diária seja menor ou igual a -10 °C.

Com base nos dados fornecidos, sua tarefa é analisar o período abrangido pelos registros e identificar quantas ondas de calor ocorreram e quando elas ocorreram (data de início da onda de calor, data de término da onda de calor e duração da onda de calor em dias).

Os dados estão em formato texto e contêm registros diários para um período especificado. Pode haver registros incompletos ou inconsistências nos dados, e você deve considerar isso ao projetar sua solução.

Bônus:

Sua solução consegue, também, detectar períodos de ondas de frio.

1. **Código-fonte:** Incluindo todos os scripts, notebooks ou arquivos necessários para processar os dados e identificar as ondas de calor.

- A solução deve ser organizada, reutilizável e bem documentada.
- Bônus: Preferencialmente, utilize as seguintes ferramentas e linguagens:
 - Python
 - Spark
 - Airflow
 - dbt

2. **Documentação:** Um arquivo README detalhando:

- A abordagem utilizada para resolver o problema.
- As ferramentas e bibliotecas utilizadas.
- Como executar sua solução.

- Cenários considerados
- Possíveis melhorias futuras.

Critérios de Avaliação

Sua solução será avaliada com base nos seguintes critérios:

1. Funcionamento da solução:

- Identificação correta das ondas de calor.
- Tratamento adequado dos dados (incluindo fluxos de error).

2. Escolha / design da solução:

- Adequação da abordagem à complexidade do problema.
- Eficiência no processamento dos dados (especialmente para grandes volumes).

3. Aderência às melhores práticas:

- Qualidade do código (organização, legibilidade, manutenibilidade).
- Uso de padrões de engenharia de software e engenharia de dados.
- Tratamento de erros e qualidade dos dados.

4. Apresentação da solução:

- No dia da apresentação, você será avaliado quanto à qualidade da exposição dos fatos e clareza e objetividade de comunicação para que a audiência entenda as suas escolhas que levaram à solução implementada.

Instruções adicionais:

- Você tem 7 dias corridos para concluir o desafio e enviar sua solução.
- Submeta o material via [e-mail].
- Caso tenha dúvidas, entre em contato pelo canal [e-mail].

Boa sorte!

Estamos ansiosos para conhecer sua solução.