

Desafio Técnico para Cientista de Dados Pleno

Objetivo

O objetivo desse desafio é avaliar sua capacidade de conduzir um projeto ponta a ponta, desde a análise exploratória até a construção de pipelines e boas práticas de MLOps. Espera-se que você construa uma solução preditiva para estimar o consumo energético de clientes residenciais, utilizando dados fornecidos em arquivos CSV. Ainda, esperamos o uso de Python, SQL com DuckDB e boas práticas de MLOps.

Dados disponíveis

Os dados estão localizados no diretório `data/`, divididos em três arquivos:

- `consumo.csv`: histórico de consumo por cliente e data.
- `clima.csv`: dados climáticos agregados por região e data (com valores faltantes).
- `clientes.csv`: dados cadastrais de clientes (com algumas regiões incorretas).

Tarefas esperadas

1. **Análise exploratória (EDA)** dos dados.
2. **Tratamento de dados ausentes e inconsistentes.**
3. **Criação de tabelas no DuckDB** com os dados normalizados.
4. **Consultas SQL** para geração de insights e features.
5. **Treinamento de pelo menos dois modelos preditivos.**
6. **Validação temporal adequada** para séries temporais.
7. **Construção de um pipeline** contendo:
 - Ingestão dos CSVs
 - Transformações
 - Treinamento
 - Inferência
8. **Registro do modelo treinado** (MLflow ou pickle versionado).
9. **Dashboard interativo** com visualizações de previsões e métricas (pode ser interativo dentro de um jupyter).
10. **Apresentação executiva** com storytelling (slides PDF ou PPT).

Tecnologias esperadas

- Python (Pandas, Scikit-learn e etc)
- DuckDB + SQL
- MLflow ou versão manual de modelos
- Git

Prazo de entrega

O prazo para entrega é de **7 dias corridos** a partir do recebimento deste teste.

Entrega

Por favor, envie um link para um repositório Git (público ou privado com acesso concedido) contendo:

```
└── data/
    ├── notebooks/
    ├── models/
    ├── dashboards/
    ├── README.md
    └── apresentacao.pdf
```

Inclua instruções claras para execução no `README.md` (ambiente, bibliotecas, ordem de execução).

Boa sorte!

Estamos ansiosos para conhecer sua solução.