DESAFIO DE CIENTISTA DE DADOS

Instruções para o desafio técnico

# Objetivo

* Neste desafio você deverá analisar, estruturar e transformar dados em formato raw fornecidos no arquivo ‘dados\_ficha\_a\_desafio.csv’
* Disponibilizamos uma breve descrição dos dados e seu contexto de geração. Essa descrição pode ser vaga para alguns campos, sendo parte do desafio a interpretação do dado.

# Descrição

Exploração e apontamento de problemas

1. Explore e analise cada um dos campos disponíveis entendendo o tipo de preenchimento que possuem.
2. Dado o contexto de ingestão e criação da informação, traga características que chamam atenção e aponte problemas contidos no dataset. A ideia aqui é avaliar a sua capacidade analítica, fique a vontade para escrever inferências sobre como o problema pode ter sido criado, questionamentos que poderíamos levar aos fornecedores ou outras análises que podem ser feitas.

Tratamento de dados com DBT

1. Crie um modelo DBT que, a partir da tabela dada e os seus apontamentos de problemas, padronize os dados criando uma tabela tratada. Pense em escrever um código modularizado e legível.

## Tecnologias Relacionadas

* Python
* SQL
* DBT

# O que iremos avaliar

* **Completude**: A solução proposta atende a todos os requisitos do desafio?
* **Simplicidade**: A solução proposta é simples e direta? É fácil de entender e trabalhar?
* **Organização**: A solução proposta é organizada e bem documentada? É fácil de navegar e encontrar o que se procura?
* **Criatividade**: A solução proposta é criativa? Apresenta uma abordagem inovadora para o problema proposto?
* **Boas práticas:** A solução proposta segue boas práticas de Python, Git, etc.?

# Atenção

* A solução desse desafio deve ser publicada em um repositório público no GitHub.
* O link do repositório deve ser enviado em até uma semana após o recebimento do desafio através do [formulário](https://forms.gle/TgrkoByBmYE9WMxu7).
* Você deverá ser capaz de apresentar sua solução, explicando como a idealizou.

# Links de Referência

* Documentação DBT: [What is dbt? | dbt Developer Hub](https://docs.getdbt.com/docs/introduction)
* Repositório de modelos DBT da SMS: [GitHub - prefeitura-rio/queries-rj-sms: Projeto dbt do Data Lake da Secretaria Municipal de Saúde](https://github.com/prefeitura-rio/queries-rj-sms/)

# Dúvidas?

* Fale com Vitoria Leite: [vitoria.leite@dados.rio](mailto:vitoria.leite@dados.rio)