# Desafio Técnico – Analytics Engineer Sênior

## 🎯 Framework de Qualidade de Dados com dbt + Great Expectations

Neste desafio, você irá desenvolver um sistema de validação de dados integrando dbt com Great Expectations, para garantir a consistência, completude e confiabilidade de um pipeline de dados analíticos.

O objetivo é estruturar um framework reutilizável que possa ser aplicado em qualquer projeto analítico com foco em governança de dados.

## 🧪 Descrição do Desafio

Você receberá (ou criará) um dataset simulando uma base de vendas ou cadastro de clientes. Seu desafio é definir e implementar testes automáticos para validar esse dataset antes da carga analítica.

Você usará dbt para transformar os dados e Great Expectations para aplicar testes de qualidade em diferentes estágios da pipeline.

## 🔧 Entregáveis Técnicos

- Projeto dbt estruturado com ao menos:

- Camada staging para limpeza

- Camada marts com modelagem final

- Integração com Great Expectations para validar:

- Nulos em campos obrigatórios

- Tipos de dados esperados

- Regras de negócio (ex: valores positivos, datas válidas)

- Relatórios de validação gerados automaticamente

- Documentação no README explicando cada teste criado

## ✨ Extras que contam pontos

- Testes customizados com SQL ou Python

- Script para rodar tudo via CLI (ex: Makefile)

- Alertas por e-mail ou log em caso de falha

- dbt docs ativo com testes visíveis

- Organização por múltiplos ambientes (dev/prod)

## ⚠ Importante

Você pode mockar os dados com CSVs ou usar um gerador de dados (Faker, Mockaroo).

O foco está na \*\*criação do framework de validação e não na beleza do dashboard final\*\*. Mostre domínio sobre consistência e validação ao longo do pipeline.

## 📦 Entrega esperada

- Repositório no GitHub com:

- Projeto dbt com estrutura clara

- Configurações e arquivos do Great Expectations

- Arquivo com logs ou prints dos testes rodando

- README.md explicando testes e como executar

## 🕒 Instruções Finais

Prazo para entrega: 3 dias corridos

Tempo estimado de execução: entre 6 a 8 horas

Nível de dificuldade: Sênior

Este desafio mede sua atenção à confiabilidade dos dados e capacidade de estruturar uma esteira robusta de qualidade.

Construa como se fosse entregar para um time global que depende dos seus dados para tomar decisões! 🚀