

Desafio Técnico – Engenheiro de Dados

O objetivo deste teste é avaliar sua capacidade de projetar e implementar um pipeline de dados utilizando boas práticas de engenharia e arquitetura, com foco em organização em camadas (arquitetura *Medallion*).

Contexto

O link abaixo disponibiliza um arquivo público da **ANS (Agência Nacional de Saúde Suplementar)** contendo informações de **operadoras e beneficiários de planos de saúde do estado de São Paulo**.

Esses dados são atualizados mensalmente e incluem informações como:

- Código e nome das operadoras
- Município de operação
- Faixa etária dos beneficiários
- Segmento do plano
- Abrangência
- Quantidade de beneficiários ativos

Fonte dos dados:

https://dadosabertos.ans.gov.br/FTP/PDA/informacoes_consolidadas_de_beneficiario_s-024/202508/pda-024-icb-TO-2025_08.zip

Tarefa

Você deverá criar um **pipeline de dados** (utilizando linguagem SQL) seguindo a **arquitetura Medallion**, estruturada em três camadas:

- 1. Bronze (Raw Layer)**
 - a. Ingestão do arquivo “as-is”, preservando a estrutura original.
- 2. Silver (Refined Layer)**
 - a. Conversão dos dados para formato tabular (ex.: Parquet ou Delta).
 - b. Aplicação de tipagem de colunas e possíveis regras de mascaramento (se necessário).
- 3. Gold (Curated Layer)**
 - a. Otimização dos dados para consumo analítico.
 - b. Aplicação de regras de performance (ex.: particionamento, agregações, caching).

Consultas solicitadas

A partir dos dados processados, responda:

- a) Quais são as **5 operadoras** com **maior número de beneficiários ativos**?
- b) Qual é a **faixa etária com mais beneficiários** e quantos são?
- c) Liste, de forma **decrescente**, a **quantidade de beneficiários por município**.

Entregáveis

- Código-fonte (SQL scripts ou notebooks).
- Breve descrição do pipeline (pode ser um README ou um diagrama simples).
- Resultados das três consultas acima.

Avaliaremos

- Clareza e organização do pipeline.
- Boas práticas de arquitetura de dados (Medallion, modularidade, performance).
- Legibilidade e estrutura do código.