

Questão 1

Para importação dos dados dos arquivos .csv, inicialmente foram criadas as tabelas de Pacientes, Exames e Desfechos, utilizando tipos de dados e *constraints* baseadas no arquivo de descrição de dados fornecido. Esse processo se deu de forma iterativa, ajustando os tipos e tamanhos conforme a importação era feita e mensagens de alerta eram geradas.

Os comandos iniciais apenas excluem as estruturas das tabelas, caso existentes, e as criam novamente com os tipos de parâmetros desejados.

Em seguida, fazemos uma configuração relacionada ao formato da data (DD/MM/AA), para que a importação dos dados do .csv ocorra com sucesso, e então ocorre a importação dos dados utilizando o comando *COPY*. Foi utilizado apenas o ***Hospital Sírio-Libanês***.

Finalmente, foi realizado o tratamento de dados. O primeiro deles está relacionado com o campo de ano de nascimento da tabela de Pacientes. Na especificação, é citado que os anos 'AAAA' são menores ou iguais a 1930, e os anos 'YYYY' seriam dados anonimizados. Assim, fizemos a substituição de todos 'AAAA' por '1930' e 'YYYY' por *null*. Então, com os anos modificados todos para tipo inteiro, alteramos a coluna de ano para *integer*.

A anonimização de dados ocorreu também em campos como unidade federativa (*cd_uf*), município (*cd_municipio*) e CEP reduzido (*cd_cepreduzido*). Em todos eles, fizemos o tratamento e alteramos para *null* tais dados anonimizados.

Em seguida, com o objetivo de padronizar e facilitar futuras buscas, removemos todos os acentos da coluna *de_analito* da tabela de Exames, além de colocar os valores em letra minúscula.

Posteriormente, foram realizados os tratamentos relacionados ao COVID, que é alvo de estudo em questões futuras. Foi realizada uma análise dos exames e analitos relacionados ao COVID, com o objetivo de padronizar as colunas de *de_analito* e *de_resultado*. Os analitos identificados ficaram com valores de 'DETECCAO COVID', 'COVID 19 - ANTICORPOS IGM' e 'COVID 19 - ANTICORPOS IGG'. Em paralelo, os resultados foram traduzidos para 'DETECTADO', 'NAO DETECTADO' e 'INCONCLUSIVO'.