

## Parte 2

### Item A)

As colunas presentes no [arquivo .csv de Alagoas](#) são:

document\_id; paciente\_id; paciente\_idade; paciente\_datanascimento;  
paciente\_enumsexobiologico; paciente\_racacor\_codigo; paciente\_racacor\_valor;  
paciente\_endereco\_coibgemunicipio; paciente\_endereco\_copais;  
paciente\_endereco\_nmmunicipio; paciente\_endereco\_nmpais;  
paciente\_endereco\_uf; paciente\_endereco\_cep;  
paciente\_nacionalidade\_enumnacionalidade; estabelecimento\_valor;  
estabelecimento\_razaosocial; estalecimento\_nofantasia;  
estabelecimento\_municipio\_codigo; estabelecimento\_municipio\_nome;  
estabelecimento\_uf; vacina\_grupoatendimento\_codigo;  
vacina\_grupoatendimento\_nome; vacina\_categoria\_codigo; vacina\_categoria\_nome;  
vacina\_lote; vacina\_fabricante\_nome; vacina\_fabricante\_referencia;  
vacina\_dataaplicacao; vacina\_descricao\_dose; vacina\_codigo; vacina\_nome;  
sistema\_origem; data\_importacao\_rnds; id\_sistema\_origem

Abaixo, detalho um pouco como pensei as tabelas:

Vacinação	
document_id	INT. <b>Chave primária.</b> Identificador único da vacinação.
paciente_id	STRING. <b>Chave estrangeira.</b> Identificador único do paciente.
vacina_dataaplicacao	DATE. Data de aplicação da vacina.
estabelecimento_valor	INT. <b>Chave estrangeira.</b> Código do estabelecimento.
vacina_grupoatendimento_codigo	SMALLINT. <b>Chave estrangeira.</b> Código do grupo de atendimento de vacinação.
vacina_categoria_codigo	SMALLINT. <b>Chave estrangeira.</b> Código da categoria do grupo de vacinação.
vacina_lote	STRING. Lote da vacina que o paciente tomou.
vacina_descricao_dose	STRING. Descrição se é 1ª ou 2ª dose.
vacina_codigo	SMALLINT. <b>Chave estrangeira.</b> Código da Vacina.
sistema_origem	STRING. Sistema de origem da vacina.
data_importacao_rnds	DATE. Data de importação do lote da vacina.

Vacina	
<code>vacina_codigo</code>	SMALLINT. <b>Chave primária.</b> Código da Vacina.
<code>vacina_nome</code>	STRING. Nome da vacina.
<code>vacina_fabricante_referencia</code>	STRING. Referência da fabricante da vacina.
<code>vacina_fabricante_nome</code>	STRING. Nome da fabricante da vacina.

Paciente	
<code>paciente_id</code>	STRING. <b>Chave primária.</b> Identificador único do paciente.
<code>paciente_idade</code>	SMALLINT. Idade do paciente.
<code>paciente_datanascimento</code>	DATE. Data de nascimento do paciente.
<code>paciente_enumsexobiologico</code>	ENUM. Sexo biológico do paciente.
<code>paciente_racacor_codigo</code>	SMALLINT. <b>Chave estrangeira.</b> Código da raça/cor.
<code>paciente_endereco_coibgemunicipio</code>	SMALLINT. <b>Chave estrangeira.</b> Código do município.
<code>paciente_nacionalidade_enumnacionalidade</code>	ENUM. Nacionalidade.
<code>paciente_endereco_cep</code>	INT. CEP do paciente.

Raça-cor	
<code>racacor_codigo</code>	SMALLINT. <b>Chave primária.</b> Código da raça/cor.
<code>paciente_racacor_valor</code>	STRING. Descrição da raça/cor.

Município	
<code>paciente_endereco_coibgemunicipio</code>	SMALLINT. <b>Chave primária.</b> Código do IBGE do município.
<code>paciente_endereco_copais</code>	SMALLINT. Código do país.

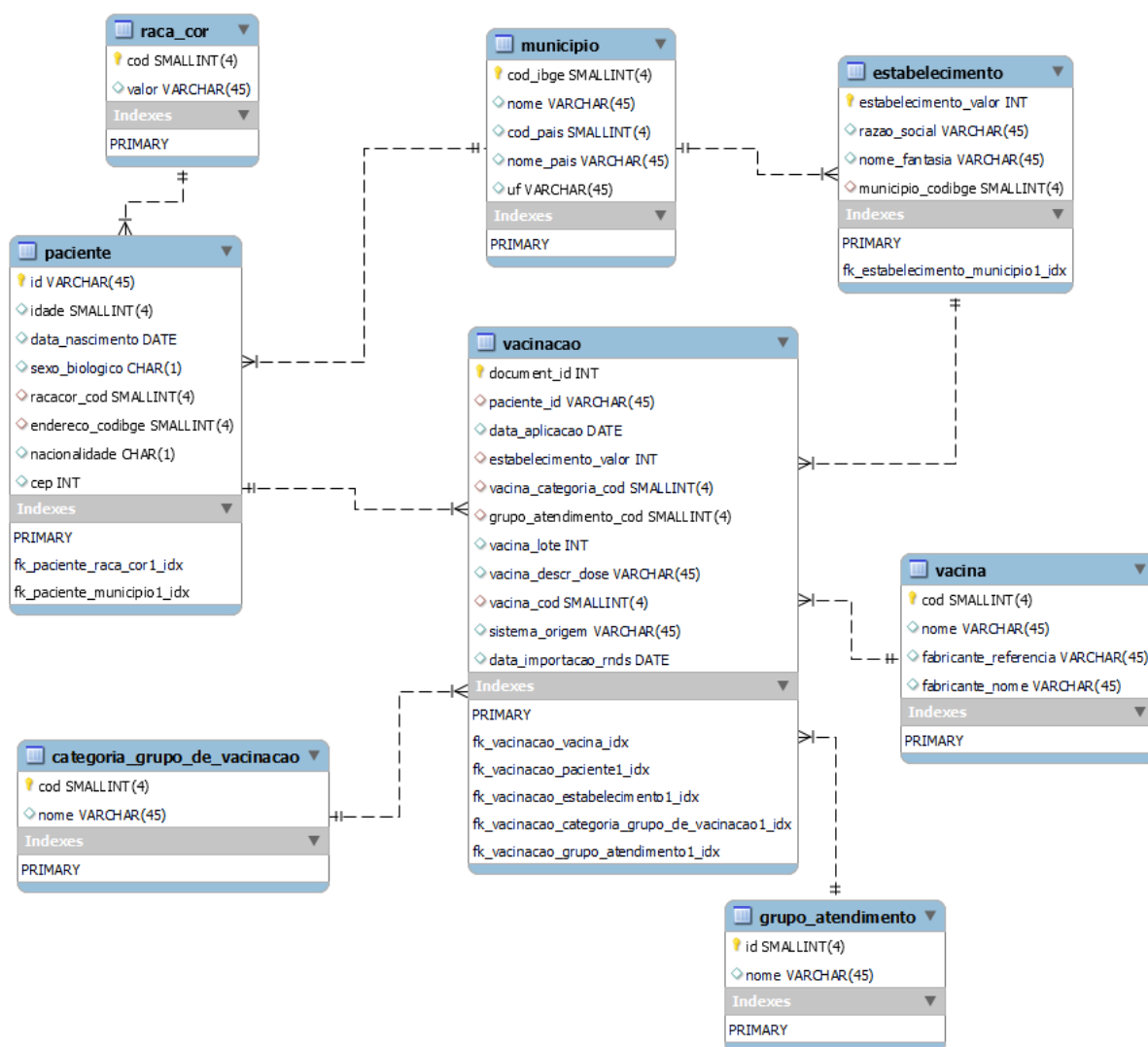
<code>paciente_endereco_nmmunicipio</code>	STRING. Nome do município.
<code>paciente_endereco_nmpais</code>	STRING. Nome do país do município.
<code>paciente_endereco_uf</code>	STRING. Unidade federativa do município.

Estabelecimento	
<code>estabelecimento_valor</code>	INT. <b>Chave primária.</b> Código do estabelecimento
<code>estabelecimento_razaosocial</code>	STRING. Razão social do estabelecimento.
<code>estabelecimento_nofantasia</code>	STRING. Nome fantasia do estabelecimento.
<code>estabelecimento_municipio_codigo</code>	SMALLINT. <b>Chave estrangeira.</b> Código do município.

Categoria grupo de vacinação	
<code>vacina_categoria_codigo</code>	SMALLINT. <b>Chave primária.</b> Código da categoria do grupo de vacinação.
<code>vacina_categoria_nome</code>	STRING. Nome da categoria de grupo de vacinação.

Grupo de atendimento	
<code>vacina_grupoatendimento_codigo</code>	SMALLINT. <b>Chave primária.</b> Código do grupo de atendimento
<code>vacina_grupoatendimento_nome</code>	STRING. Nome do grupo de atendimento

Um diagrama ER (com alguns nomes de colunas simplificados em relação ao que tínhamos no .csv) pode ser visualizado na imagem abaixo:



A imagem acima pode ser visualizada em resolução melhor no repositório do github:

[https://github.com/victoraccete/LED\\_tasks/blob/main/led\\_task\\_pt2.png](https://github.com/victoraccete/LED_tasks/blob/main/led_task_pt2.png)

Usando a funcionalidade de forward engineering do MySQL Workbench, obtivemos os comando para implementar o banco. O código está disponível em:

[https://github.com/victoraccete/LED\\_tasks/blob/main/led\\_task\\_pt\\_2\\_forward\\_engineer.sql](https://github.com/victoraccete/LED_tasks/blob/main/led_task_pt_2_forward_engineer.sql)

## Item B)

O notebook do item B está disponível [aqui \(github\)](#).

## Item C) Discussão.

Acredito que há problemas de anonimização dos dados, principalmente devido a dados excessivos, o que é uma possível desconformidade com a LGPD. É verdade que não há dados muito pessoais, como “Nome”, “CPF” ou “Documento de identidade”. No entanto, a combinação de alguns dados como “sexo biológico”, “data de nascimento” e “cep” podem fazer com que pessoas mal intencionadas saibam de quem se trata.

Sabemos que a combinação de alguns dados pode fazer com que a privacidade seja quebrada, mesmo que o dataset em si não seja revelador. Um [caso famoso](#) foi o de uma competição da Netflix:

*“In 2007 two researchers from The University of Texas at Austin were able to identify individual users by matching the data sets with film ratings on the Internet Movie Database.”*

Ou seja, apesar dos dados estarem anonimizados, de certa forma, e o próprio CEP não estar completo, ainda pode ser um pouco problemático se cruzado com informações de outras bases.

Acredito que uma proposta de solução seria incluir apenas o ano de nascimento, ou mês e ano, sem o dia, para dificultar o cruzamento com outras bases. Ocultar o CEP completamente prejudicaria muito a transparência, pois inviabilizaria análises de vacinação por bairro, por exemplo.