

Proposta de Desafio para a equipe de BI

Introdução

Suponha que você tenha um sistema com as seguintes características:

Pessoas são cadastradas no sistema e possuem as informações de nome, data de nascimento, CPF, sexo, raça e informação se é pessoa com deficiência (sim/não).

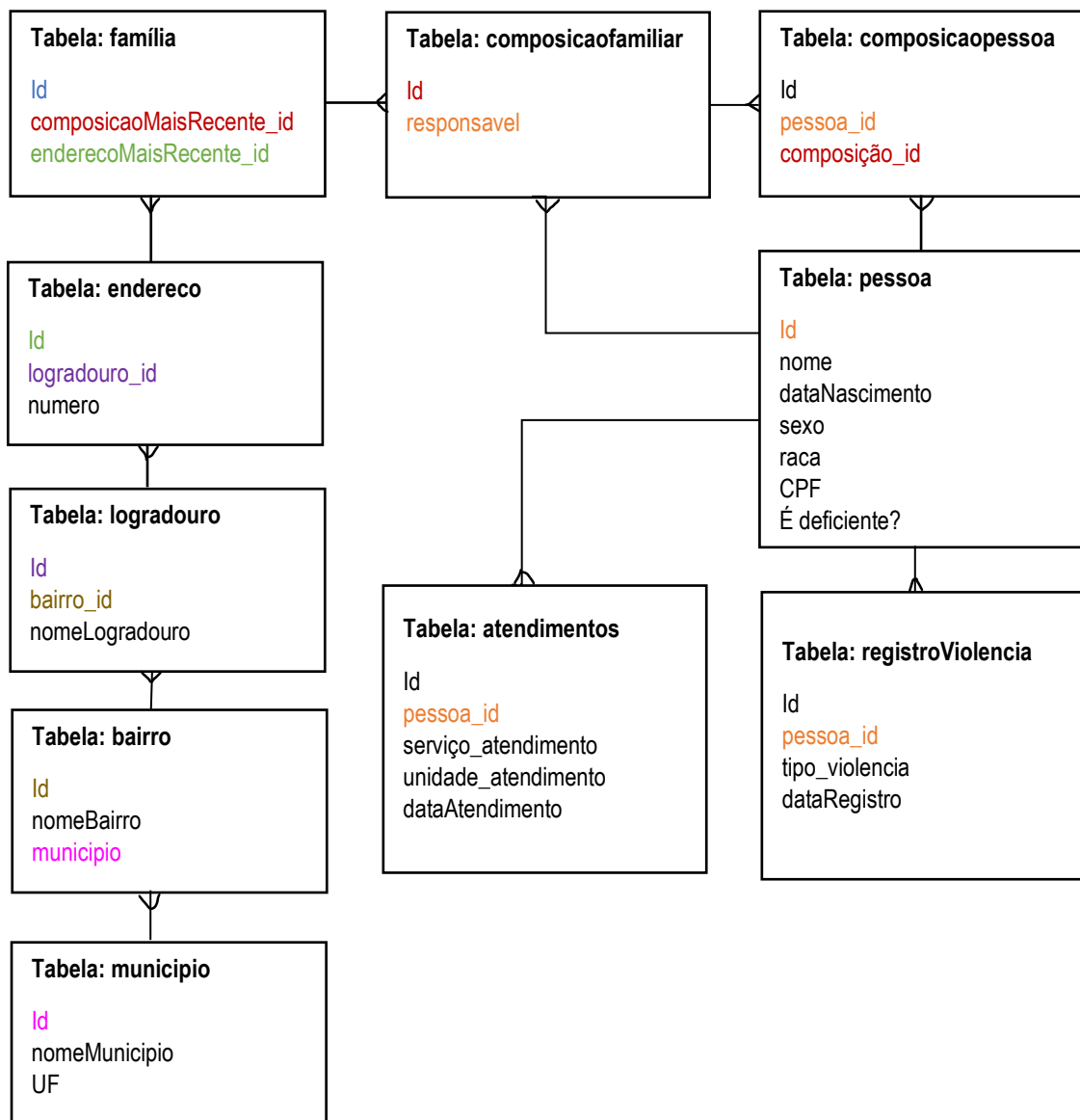
Famílias são cadastradas no sistema e possuem as seguintes informações: identificação do responsável pela família, endereço (logradouro, número, bairro, cidade e estado) e identificação da composição familiar (ou seja, quem são as pessoas dessa família).

Sobre a composição familiar: uma família pode ter várias composições familiares, porque ao longo do tempo pessoas podem entrar na família (nascimentos, casamentos...) e podem sair da família (óbito, um filho que se casou e foi formar uma nova família...). No nosso cenário, não vamos considerar o histórico das composições familiares, para simplificar. Vamos considerar somente o relacionamento da família com a composição mais recente dela.

Dentro dos municípios que usam esse sistema, existem várias unidades de atendimento. Pessoas são atendidas nessas unidades e são armazenadas as seguintes informações: identificação da pessoa atendida, data do atendimento, unidade que prestou o atendimento e serviço socioassistencial referente ao atendimento.

No sistema, existe ainda um cadastro das violências sofridas por pessoas. Essas violências são pré-determinadas: violência física, violência psicológica, violência sexual ou abandono. É possível registrar a identificação da pessoa, o tipo de violência sofrida e a data do registro.

O diagrama abaixo apresenta o modelo relacional simplificado do banco de dados deste sistema.



1) Modelo Dimensional

Como você organizaria esses dados em um modelo dimensional? Deixe claro quais são as dimensões e quais são as fatos.

2) Consultas SQL

Escreva consultas SQL para gerar os seguintes resultados a partir da base de dados relacional do diagrama apresentado na introdução:

- a) Gerar uma listagem das famílias cadastradas, contendo as seguintes informações:
 - Nome do responsável, CPF do responsável, bairro de residência e quantidade de pessoas na família.
- b) Gerar uma listagem com a quantidade de pessoas que sofrem violência sexual, agrupando pelas seguintes faixas etárias: crianças (0 a 12 anos), adolescentes (13 a 17 anos), adultos (18 a 59 anos) e idosos (60 anos ou mais)
- c) Gerar uma listagem de bairros com o total de famílias e o total de pessoas por bairro.
- d) Gerar uma listagem dos tipos de violência e o número de pessoas que tiveram registro para cada uma delas, no mês de janeiro de 2023.

3) Python

Crie um código em Python (preferencialmente dentro de um Jupyter Notebook, caso conheça esse ambiente), que gere um arquivo csv de 100 linhas, com os seguintes campos:

- idPessoa: número sequencial de 1 a 100
- nome: imprimir nomes sequenciais como Nome1, Nome2, Nome3... até Nome100
- dataNascimento: criar datas quaisquer para simular datas de nascimento. Mas todas devem ser anteriores a 2020 (sugiro pensar numa forma que gere pessoas com idades diversificadas).
- indDeficiencia: preencher essa coluna aleatoriamente com os valores 0 ou 1, pra indicar se a pessoa tem ou não alguma deficiência.

Este arquivo deve ter o nome dPessoas.csv

Crie um outro código em Python que gere um outro arquivo csv, este com 200 linhas, contendo os seguintes campos:

- idAtendimento: número sequencial de 1 a 200
- idPessoa: números aleatórios entre 1 e 100 (para simular as pessoas geradas anteriormente)
- dataAtendimento: criar datas quaisquer para simular datas de atendimento. Podem ser sequenciais ou aleatórias. Mas todas devem ser do ano de 2023 (sugiro ter registro em todos os meses do ano).

Este arquivo deve ter o nome fAtendimentos.csv

4) Power BI

Criar um projeto no Power BI Desktop com o nome DesafioBIGesuas.pbix

No Power Query:

Importar os dois arquivos criados anteriormente (dPessoas.csv e fAtendimentos.csv). Importar os arquivos enviados dViolencias.csv e fRegistroViolencia.csv. A tabela fato de violências contém o id da pessoa que sofreu a violência e o id do tipo da violência.

Formatar os campos.

Criar na tabela dPessoas uma nova coluna com a idade atual da pessoa.

Criar na tabela dPessoas uma nova coluna com a faixa etária da pessoa: crianças (0 a 12 anos), adolescentes (13 a 17 anos), adultos (18 a 59 anos) e idosos (60 anos ou mais)

Criar uma tabela Calendário contendo todas as datas que vão desde a primeira data de atendimento até a última data de atendimento da tabela fAtendimentos.

No Power BI:

Conectar as tabelas no modelo de dados do Power BI através dos campos idPessoa. A tabela calendário deve se ligar somente na tabela de Atendimentos.

Criar um visual que mostre o número de pessoas atendidas por faixa etária.

Criar um visual que mostre o número de pessoas com deficiência que foram atendidas.

Criar um visual que mostre uma lista com o nome e o indicador de pessoa com deficiência das pessoas atendidas.

Criar um gráfico que mostre a quantidade de atendimentos na linha do tempo, por mês, comparando com o mesmo período do mês anterior.

Criar um visual que mostre a quantidade de violências registradas, por tipo de violência.

Criar um filtro que possibilite filtrar todos os dados da página por período (data)