

Exploração e apontamento de problemas

Problemas mais gerais:

- **Colunas com valores booleanos:** necessário padronizar pois contém tanto True e False quanto 0 e 1.
- **Geral:** Inconsistência provavelmente ocorrendo na codificação de caracteres pois palavras com acento vem desformatadas, além de haver listas, json array... Pode ser uma sugestão de ajuste feito no sistema.
- **Colunas com data-hora:** parte das colunas está com mais de um formato de data e hora. Necessário padronizar tudo para ex. DD/MM/AAAA HH:MM. Mais uma sugestão para ajustar no sistema ao se inserir a data e hora.

Problemas mais específicos:

- **id_paciente** parece ser um identificador único no formato de **UUID**, apesar disso, possui campos duplicados e triplicados. Necessário fazer que os ID sejam únicos para cada pessoa (provavelmente será PK da tabela).
- **obito:** booleano que varia entre 0,1, true ou false - inconsistência. Será necessário padronizar e uma sugestão seria pedir ao fornecedor do sistema que permita a inserção apenas de valores pré definidos.
- **bairro:** Existem valores que parecem corresponder ao mesmo bairro porém escritos de maneira diferente. Para não haver tal problema, o sistema pode possui uma lista com os nomes dos bairros para serem selecionados.
- **raca_cor; bairro, identidade_genero:** valores como 'não' e 'sim' que podem trazer interpretações ambíguas. Acredito que seja melhor especificar mais, por ex. 'Não quis responder'. Sim não faz sentido. Bom possuir listas pré-definidas no sistema.
- **ocupacao:** Inconsistência na padronização dos dados inseridos. Ex Advogados com especialidades definidas possuem a especialidade entre parênteses, enquanto Advogado empresarial não está seguindo esse formato, podendo ser Advogado (direito empresarial). A mesma coisa ocorre para outras profissões, caso faça sentido para a organização, algumas profissões poderiam ser unidas em um valor, exemplo as variações de vendedor. além de outras inconsistências como nomes com as iniciais das palavra em maiusculo e em minúsculo, formato inconsistente como nas colunas anteriores. Seria melhor ter uma lista pré definida no sistema por onde os dados são coletados.

- **religiao, situacao_profissiona, renda_familiarl** : alguns campos estão preenchidos com valores que não correspondem ao que pede a coluna, podendo ter sido entendido de maneira incorreta a pergunta pelo entrevistado e/ou entrevistador. O sistema poderia oferecer uma lista pré definida de valores.
- **meios_transporte**: Existe tanto “alternativo”, quando meios também popularmente conhecidos como alternativos, como bicicleta, seria melhor que o sistema possuísse uma lista de meios de transporte para não haver ambiguidade e incerteza sobre ao que está se referindo.