**TP2 INTRODUÇÃO A BANCO DE DADOS**

**1-** Escolher dois ou mais tipos de fontes de dados em alguma fonte publica

a. Descrição geral dos acidentes de trânsito com vítima em 2022:

* <https://dados.pbh.gov.br/dataset/relacao-de-ocorrencias-de-acidentes-de-transito-com-vitima/resource/1e3d3bc4-7256-4f36-8419-12c28f250869>

b. Descrição das pessoas envolvidos nos acidentes de trânsito com vítima em 2022:

* <https://dados.pbh.gov.br/dataset/relacao-das-pessoas-envolvidas-nos-acidentes-de-transito-com-vitima/resource/8ee6609b-82de-4b93-8051-2f9d1e332784>

**2 -** Carregar esses dados em um gerenciador de bancos de dados relacional, como o PostgreSQL ou outro, e combinar esses dados, de modo a permitir análises integradas

a. Os dois conjuntos mostram aspectos diferentes de um mesmo problema no trânsito (acidentes).

b. Uso de cinto de segurança e álcool influenciam em diferentes tipos de acidentes e número de acidentes com ferimentos ou fatalidades.

**3 –** Realizar uma análise exploratória nos dados

a.

b. Descrição das pessoas envolvidos / Descrição geral dos acidentes:



c. Descrição geral dos acidentes de trânsito com vítima em 2022:

* Fonte: <https://dados.pbh.gov.br/dataset/relacao-das-pessoas-envolvidas-nos-acidentes-de-transito-com-vitima/resource/8ee6609b-82de-4b93-8051-2f9d1e332784>
* Data de obtenção: 13/06/2024
* Órgão produtor: Prefeitura de Belo Horizonte
* Data de referência: 31 de agosto de 2023
* Limitações registradas: nenhuma
* Cobertura: cidade de Belo Horizonte

Descrição das pessoas envolvidos nos acidentes de trânsito com vítima em 2022:

* Fonte: <https://dados.pbh.gov.br/dataset/relacao-de-ocorrencias-de-acidentes-de-transito-com-vitima/resource/1e3d3bc4-7256-4f36-8419-12c28f250869>
* Data de obtenção: 13/06/2024
* Órgão produtor: Prefeitura de Belo Horizonte
* Data de referência: 31 de agosto de 2023
* Limitações registradas: nenhuma
* Cobertura: cidade de Belo Horizonte

**4 -** Apresentar uma análise crítica das fontes de dados utilizadas

* Há muitos nulos ou ‘NÃO INFORMADOS’ nas tabelas, principalmente na DESC\_GERAL e em uma coluna crucial para a análise, a coluna ‘embreagues’ (na tabela original a coluna conta com esse erro de português).
* A atualização é muito tardia e conta com certa imprecisão quando a localidade e coordenadas de GPS.
* Há atributos que possuem o caractere de virgula ‘,’ em sua composição, na extração dos dados do site em CSV, que também é separado por virgulas, há conflito quando se vai importar para o banco de dados PostgreSQL.
* Há também a possibilidade de que a vitima do acidente possa ter mentido informações como se usavam cinto de segurança ou não em casos em que não havia possibilidade de usar um método de verificação mais preciso.

**5 -** Apresentar análises referentes à combinação ou integração de dados.

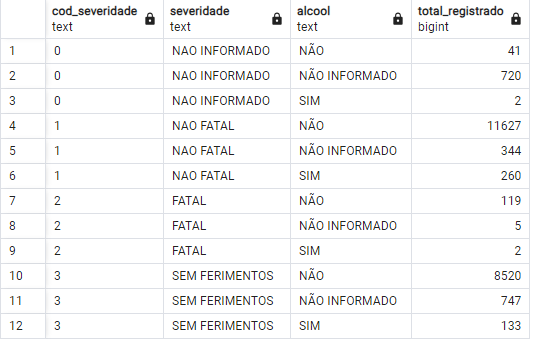
* Todos os dados vem de boletins de ocorrência registrados pela polícia, logo é possível correlacionar as bases pelo numero do boletim de ocorrência e com isso realizar o cruzamento de dados para que se tirem as analises previamente citadas.

**1** – *Preparação dos dados*: Em ambas as bases de dados foi necessário corrigir dados em que haviam o caractere virgula ‘,’ em sua composição, pois também era o delimitador do arquivo CSV, o que gerava erro ao se importar a tabela para o PostgreSQL. Para resolver isso, usei um script em python para colocar todas as informações das colunas entre aspas duplas e assim, o SGBD reconheceu somente as virgulas certas como delimitadores. (O script também está no repositório do GITHUB)

Além disso, na tabela DESC\_GERAL, houve um erro na criação coluna ‘DATA HORA\_BOLETIM’, pois, por estar separado por espaço, a palavra DATA estava sendo reconhecida como declaração de tipagem. Foi corrigido mudando o nome da coluna para ‘DATA\_HORA\_BOLETIM’.

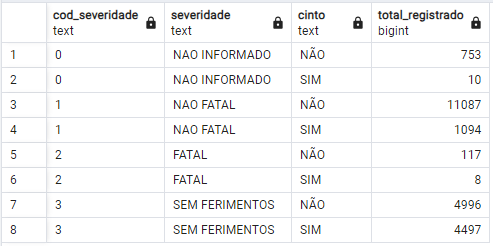
**2** – *Definição de objetivos:* O objetivo da análise é verificar a influência que embriaguez e uso de cinto de segurança causam na gravidade e frequência de acidentes nos quais as pessoas envolvidas estavam embriagadas e/ou não usavam cinto de segurança, tanto motorista quanto passageiros ou em casos de pedestres envolvidos, como esses fatores também influenciam.

**3** – Análise descritiva, identificação de valores discrepantes, análise de correlação: Durante a análise dos dados foi possível observar que a variável de embriagues tem muitas aparições como: “NÃO INFORMADO”, o que cria um viés na análise, pois torna-se inconclusivo se e quanto a embriagues do motorista influencia na fatalidade que os acidentes tiveram.

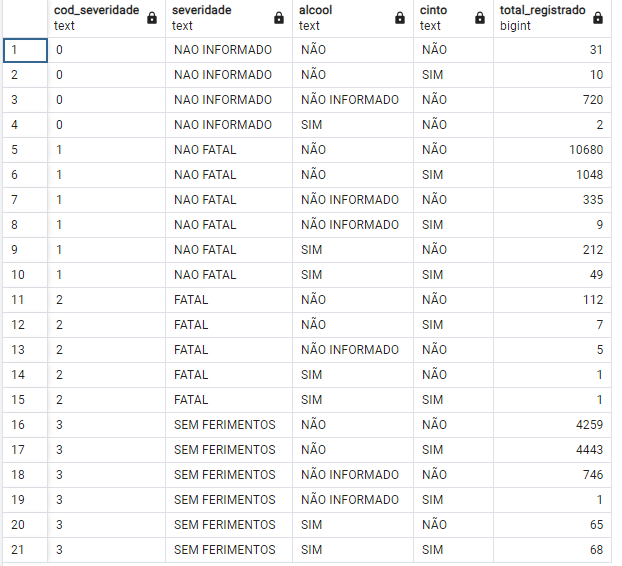


Na consulta, pode-se observar que há mais não informados do que confirmados em todos os casos.

Outra coisa a se observar é a influência do uso do cinto de segurança, independente do tipo de acidente, a severidade do acidente é grandemente influenciada pelo uso ou não do cinto de segurança. Em todos os tipos de severidade listados, sem distinção do tipo de acidente, o número de vítimas que não usavam cinto de segurança é mais de 10 vezes maior do que as vítimas que usaram, chegando até a 75,3 vezes maior no caso de severidade não informada.



Infelizmente pelo alto número de inconclusões de constatação de embriagues a analise feita juntamente da variável de cinto de segurança, fica inacabada e enviesada, como se vê a seguir:



4. Conclusões finais: Após analisar os dados fica evidente que o uso de cinto de segurança diminui a possível gravidade de um acidente com vítima, porém, fica pouco preciso e não é possível cravar que embriagues é um fator que, ao contrario do cinto de segurança, aumenta a possível gravidade de um acidente, independente, de seu tipo.

Os dados utilizados encontram-se nesse repositório público no GitHub:

<https://github.com/erickess2002/tp2_ibd.git>