

Trabalho Prático - Introdução a Banco de Dados

Relatório / Análise Crítica

Estudantes:

- Iago Diniz Martins;
- João Gabriel Macedo

Fontes dos dados:

- Sistema de Boletins de Ocorrência de Acidentes de Trânsito (Prefeitura de Belo Horizonte);
 - Sistema de Informações Ambientais (Prefeitura de Belo Horizonte);
 - Sistema de Logística/Infraestrutura (Prefeitura de Belo Horizonte);
-

Análise qualitativa dos dados e das fontes

Pontos positivos:

- Contém o ID do acidente, facilitando a integração com as demais bases (utilização com chave primária/estrangeira);
- Possui uma cobertura relativamente boa de informações como data, horário, tipo de acidente, regional e indicador de fatalidade;
- Apresenta boa cobertura para o município de Belo Horizonte, comparativamente com outros locais (como Rio de Janeiro, Salvador e Recife);
- Disponibiliza uma variedade/profundidade interessante de dados, como tipo de envolvimento (condutor, passageiro, pedestre), sexo, severidade, uso de cinto e indício de embriaguez, mesmo que com algumas irregularidades, permitindo análises mais detalhadas sobre o perfil das vítimas e dos demais envolvidos;
- Permite identificar o tipo de veículo envolvido no acidente, possibilitando observar padrões acerca desse aspecto;
- Viabiliza análises sobre a frequência de tipos de veículos em acidentes fatais, gerando indícios acerca da segurança de determinados tipos de veículos;
- Possibilita identificar a região do acidente, permitindo análises acerca da segurança do tráfego em cada localidade;
- Auxilia na identificação de áreas com maior incidência de acidentes;

Limitações:

- Existência de muitos valores nulos em campos relevantes, como tipo de acidente e indicador de fatalidade;
- Algumas categorias de tipos de acidente apresentam nomenclaturas fora de padrão;
- Não há dados de todos os anos de forma consistente, não permitindo análises que avaliem a evolução dos acidentes a longo prazo (por isso optamos por limitar os dados a 2024);
- Não há informações detalhadas sobre condições climáticas no momento do acidente (atributo que existe nas tabelas, mas cujo preenchimento é completamente inconsistente), o que limita análises ambientais mais profundas.

- Alta quantidade de valores nulos nos campos referentes ao uso de cinto de segurança e embriaguez;
- Aparente subnotificação de informações sensíveis, como consumo de álcool;
- Em alguns registros, a severidade da condição não se encontra claramente especificada;
- Não há informações sobre idade das vítimas em todos os registros.
- A elevada quantidade de valores ausentes reduz a confiabilidade das correlações entre severidade, uso de cinto e embriaguez, podendo subestimar relações importantes.
- Falta de padronização nas categorias de veículos (por exemplo, variações de nomenclatura para motocicletas);
- Ausência de informações mais detalhadas sobre estado do veículo ou ano de fabricação;
- Alguns registros apresentam dados incompletos sobre o veículo;
- Inconsistências na escrita de nomes de logradouros;
- Alguns registros indicam apenas a regional, sem detalhamento da via.

Considerações:

De forma geral, os dados disponibilizados pela Prefeitura de Belo Horizonte permitiram que realizássemos análises importantes sobre segurança no trânsito. Entretanto, há alguns erros e inconsistências recorrentes nas fontes, como:

- Elevada presença de valores nulos em atributos importantes;
- Falta de padronização na forma de escrever/apresentar os dados;
- Possível subnotificação de informações sensíveis, como embriaguez;
- Ausência de dados consistentes de outros anos, para análises de evolução através dos anos;

Tais fatores impactam negativamente na fidelidade e profundidade das análises realizadas.

Análise dos resultados

Número médio de pessoas envolvidas nos acidentes e que são ou pedestres ou condutores ou passageiros:

- Código SQL da consulta:

Nº médio de pessoas envolvidas nos acidentes e que são ou pedestres ou condutores ou passageiros: média: 2,2014 (aproximadamente 2 pessoas)

```
WITH envolvidos_por_boletim AS (
    SELECT
        No_boletim,
        COUNT(*) AS total_envolvidos
    FROM si_env_2024
    WHERE condutor = 'S'
        OR passageiro = 'S'
        OR pedestre = 'S'
    GROUP BY No_boletim
)
SELECT AVG(total_envolvidos) AS media_envolvidos_por_boletim
FROM envolvidos_por_boletim
```

Número médio de veículos envolvidos em acidentes:

- Código SQL da consulta:

```
# N° médio de veículos envolvidos em acidentes: 1,8725
```

```
WITH veiculos_por_boletim AS (  
    SELECT  
        No_boletim,  
        COUNT(*) AS total_veiculos  
    FROM si_veic_2024  
    GROUP BY No_boletim  
)  
SELECT  
    AVG(total_veiculos) AS media_veiculos_por_boletim  
FROM veiculos_por_boletim;
```

Número de acidentes por número de veículos:

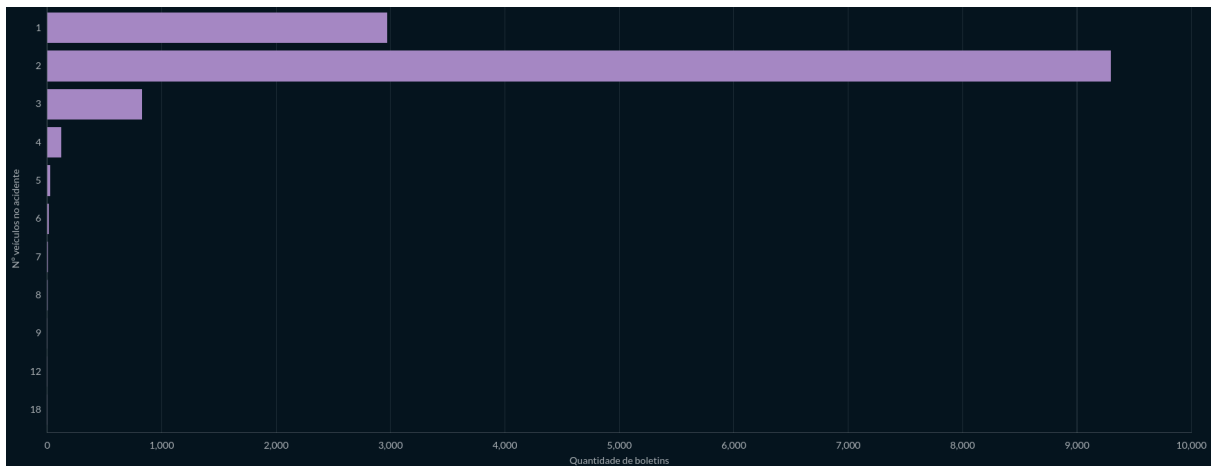
- Código SQL da consulta:

```
SELECT total_veiculos, COUNT(*) AS qtd_boletins  
FROM (  
    SELECT  
        No_boletim,  
        COUNT(*) AS total_veiculos  
    FROM si_veic_2024  
    GROUP BY No_boletim  
) x  
GROUP BY total_veiculos  
ORDER BY total_veiculos;
```

- Tabela:

total_veiculos	qtd_boletins
1	2,970
2	9,295
3	827
4	125
5	27
6	11
7	5
8	3
9	1
12	1
18	1

- Gráfico:



Quais categorias de veículos aparecem com mais frequência em acidentes fatais:

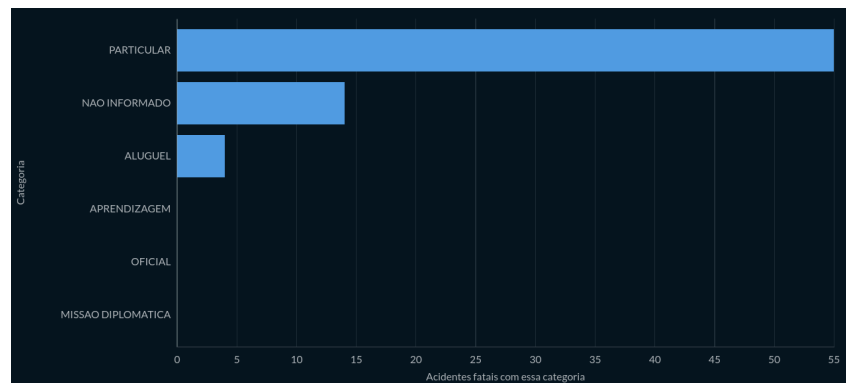
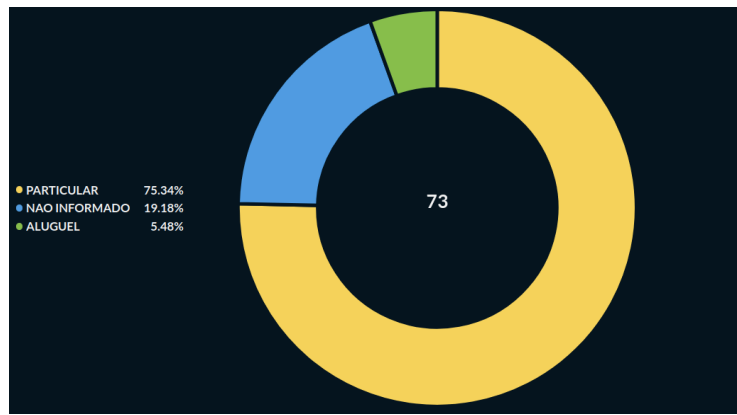
- Código SQL da consulta:

```
WITH sev AS (  
    SELECT  
        No_boletim,  
        MAX(cod_severidade) AS sev  
    FROM si_env_2024  
    WHERE condutor = 'S'  
    GROUP BY No_boletim  
)  
SELECT  
    v.descricao_categoria,  
    SUM(CASE WHEN s.sev = 2 THEN 1 ELSE 0 END) AS  
    acidentes_fatais_com_essa_categoria,  
    COUNT(*) AS total_acidentes_com_essa_categoria  
FROM si_veic_2024 v  
JOIN sev s USING (No_boletim)  
WHERE v.descricao_categoria IS NOT NULL AND v.descricao_categoria != ''  
GROUP BY v.descricao_categoria  
ORDER BY acidentes_fatais_com_essa_categoria DESC;
```

- Tabela:

descricao_categoria	acidentes_fatais_com_essa_categoria	total_acidentes_com_essa_categoria
PARTICULAR	55	21,086
NAO INFORMADO	14	1,772
ALUGUEL	4	1,783
APRENDIZAGEM	0	39
OFICIAL	0	133
MISSAO DIPLOMATICA	0	9

- Gráficos:



Qual a distribuição de acidentes por região e seu indicador de fatalidade:

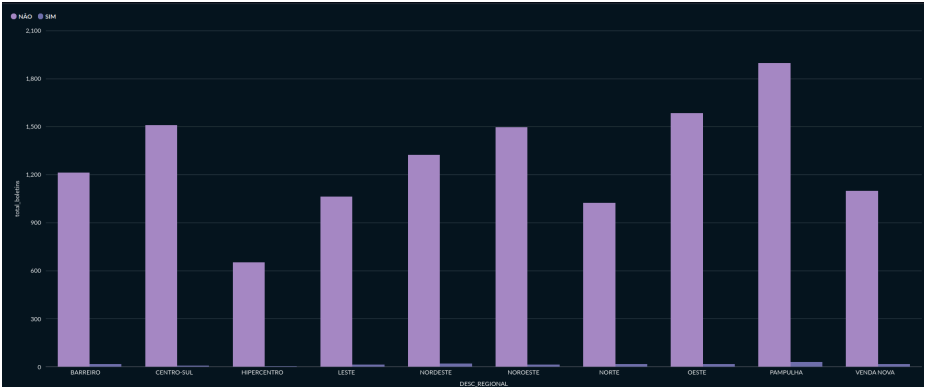
- Código SQL da consulta:

```
SELECT
    DESC_REGIONAL,
    INDICADOR_FATALIDADE,
    COUNT(DISTINCT NUMERO_BOLETIM) AS total_boletins
FROM si_bol_2024
WHERE DESC_REGIONAL IS NOT NULL AND DESC_REGIONAL != ''
    AND INDICADOR_FATALIDADE IS NOT NULL AND INDICADOR_FATALIDADE != ''
GROUP BY DESC_REGIONAL, INDICADOR_FATALIDADE
ORDER BY DESC_REGIONAL, INDICADOR_FATALIDADE;
```

- Tabela:

DESC_REGIONAL	INDICADOR_FATALIDADE	total_boletins
BARREIRO	NÃO	1,214
BARREIRO	SIM	16
CENTRO-SUL	NÃO	1,508
CENTRO-SUL	SIM	8
HIPERCENTRO	NÃO	652
HIPERCENTRO	SIM	3
LESTE	NÃO	1,063
LESTE	SIM	13
NORDESTE	NÃO	1,325
NORDESTE	SIM	19
NOROESTE	NÃO	1,497
NOROESTE	SIM	12
NORTE	NÃO	1,023
NORTE	SIM	15
OESTE	NÃO	1,583
OESTE	SIM	18
PAMPULHA	NÃO	1,897
PAMPULHA	SIM	28
VENDA NOVA	NÃO	1,098
VENDA NOVA	SIM	16

- Gráfico:



Qual a distribuição de acidentes por tipo de acidente da tabela “si_bol_2024”:

- Código SQL da consulta:

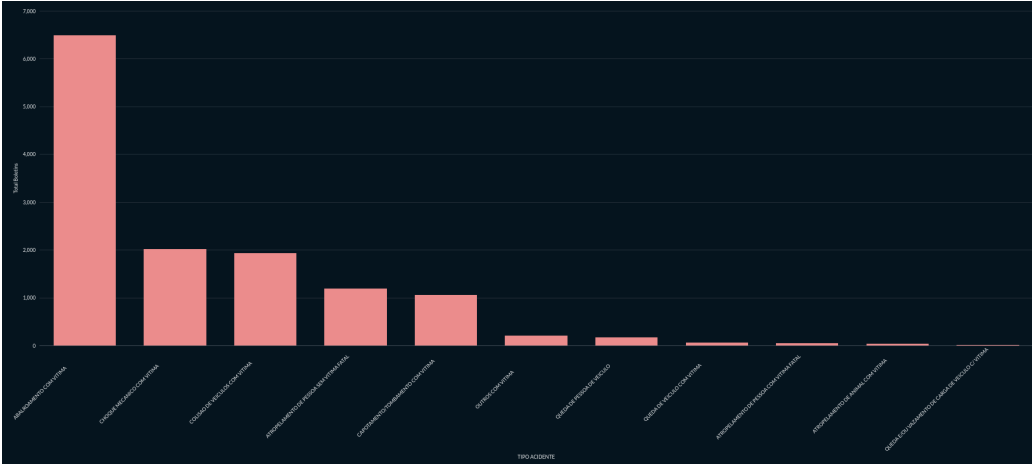
```
WITH severidade_boletim AS (  
    SELECT  
        No_boletim,  
        MAX(cod_severidade) AS cod_severidade  
    FROM si_env_2024  
    WHERE cod_severidade IN (1,2,3)  
    GROUP BY No_boletim  
)
```

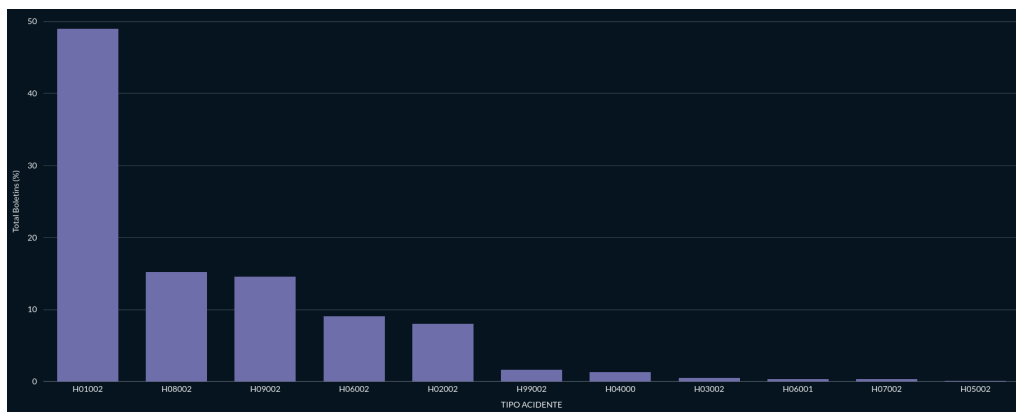
```
SELECT
    b.PAVIMENTO,
    s.cod_severidade,
    CASE s.cod_severidade
        WHEN 1 THEN 'NAO FATAL'
        WHEN 2 THEN 'FATAL'
        WHEN 3 THEN 'SEM FERIMENTOS'
    END AS desc_severidade,
    COUNT(*) AS total_acidentes
FROM severidade_boletim s
JOIN si_bol_2024 b
    ON b.NUMERO_BOLETIM = s.No_boletim
WHERE b.PAVIMENTO IS NOT NULL AND b.PAVIMENTO != ''
GROUP BY
    b.PAVIMENTO,
    s.cod_severidade
ORDER BY
    b.PAVIMENTO,
    s.cod_severidade;
```

- Tabela:

TIPO_ACIDENTE	DESC_TIPO_ACIDENTE	total_boletins	porcentagem
H01002	ABALROAMENTO COM VITIMA	6,493	48.94
H08002	CHOQUE MECANICO COM VITIMA	2,016	15.2
H09002	COLISAO DE VEICULOS COM VITIMA	1,936	14.59
H06002	ATROPELAMENTO DE PESSOA SEM VITIMA FATAL	1,199	9.04
H02002	CAPOTAMENTO/TOMBAMENTO COM VITIMA	1,064	8.02
H99002	OUTROS COM VITIMA	212	1.6
H04000	QUEDA DE PESSOA DE VEICULO	170	1.28
H03002	QUEDA DE VEICULO COM VITIMA	69	0.52
H06001	ATROPELAMENTO DE PESSOA COM VITIMA FATAL	51	0.38
H07002	ATROPELAMENTO DE ANIMAL COM VITIMA	41	0.31
H05002	QUEDA E/OU VAZAMENTO DE CARGA DE VEICULO C...	15	0.11

- Gráfico:





Relação da severidade de condutores/passageiros com a embriaguez:

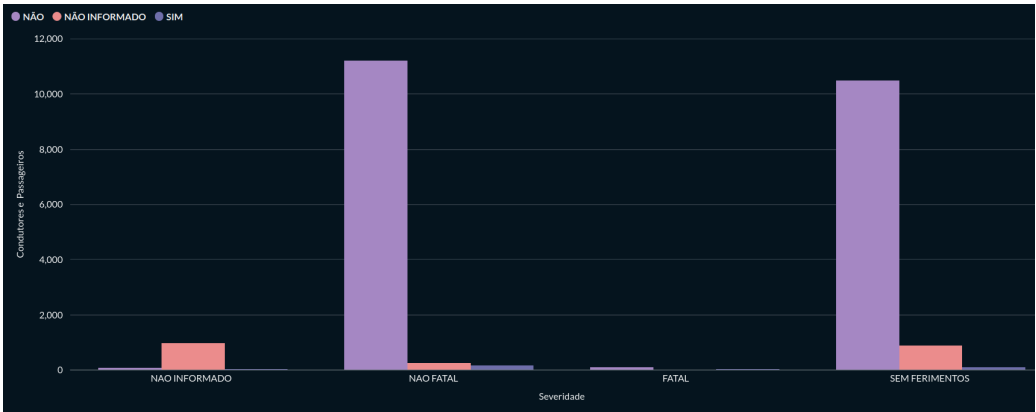
- Código SQL da consulta:

```
-- CUIDADOS:
-- 1) Considerar apenas boletins válidos (excluindo registros com
TIPO_ACIDENTE nulo ou vazio).
-- 2) Contar boletins únicos, evitando duplicidade se houver múltiplos
registros relacionados a um boletim.
SELECT
    TIPO_ACIDENTE,
    DESC_TIPO_ACIDENTE,
    COUNT(DISTINCT NUMERO_BOLETIM) AS total_boletins,
    ROUND(100 * COUNT(DISTINCT NUMERO_BOLETIM) /
        (SELECT COUNT(*)
         FROM si_bol_2024
         WHERE TIPO_ACIDENTE IS NOT NULL AND TIPO_ACIDENTE != ''), 2)
AS porcentagem
FROM si_bol_2024
WHERE TIPO_ACIDENTE IS NOT NULL AND TIPO_ACIDENTE != ''
GROUP BY TIPO_ACIDENTE, DESC_TIPO_ACIDENTE
ORDER BY total_boletins DESC;
```

- Tabela:

cod_severidade	desc_severidade	embreagues	total
0	NAO INFORMADO	NÃO	69
0	NAO INFORMADO	NÃO INFORMADO	959
0	NAO INFORMADO	SIM	13
1	NAO FATAL	NÃO	11,209
1	NAO FATAL	NÃO INFORMADO	253
1	NAO FATAL	SIM	153
2	FATAL	NÃO	93
2	FATAL	SIM	1
3	SEM FERIMENTOS	NÃO	10,489
3	SEM FERIMENTOS	NÃO INFORMADO	889
3	SEM FERIMENTOS	SIM	106

- Gráfico:



Relação da severidade de condutores/passageiros com o uso de cinto de segurança:

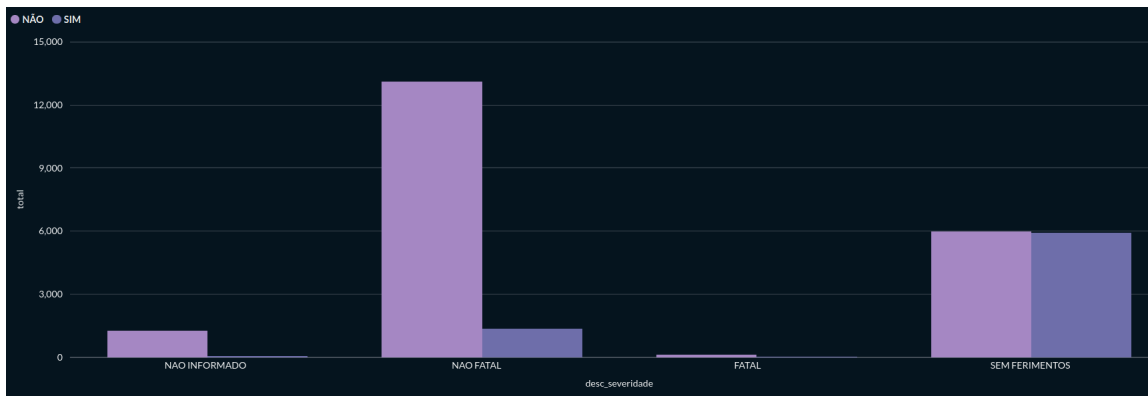
- Código SQL da consulta:

```
SELECT
    cod_severidade,
    desc_severidade,
    embreagues,
    COUNT(*) AS total
FROM si_env_2024
WHERE condutor = 'S'
GROUP BY
    cod_severidade,
    desc_severidade,
    embreagues
ORDER BY
    cod_severidade,
    embreagues;
```

- Tabela:

cod_severidade	desc_severidade	cinto_seguranca	total
0	NAO INFORMADO	NÃO	1,246
0	NAO INFORMADO	SIM	64
1	NAO FATAL	NÃO	13,102
1	NAO FATAL	SIM	1,348
2	FATAL	NÃO	105
2	FATAL	SIM	4
3	SEM FERIMENTOS	NÃO	5,983
3	SEM FERIMENTOS	SIM	5,925

- Gráfico:



Relação dos sexos dos condutores em acidentes envolvendo apenas 2 condutores:

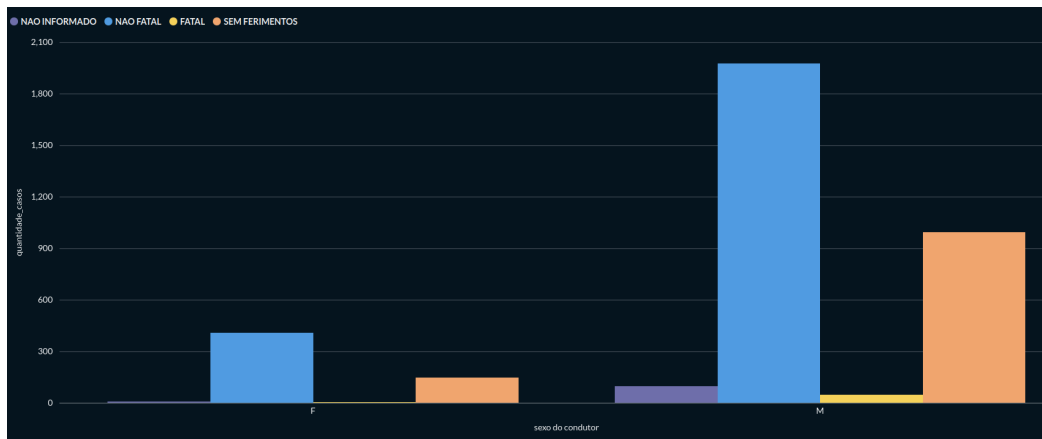
- Código SQL da consulta:

```
SELECT
    cod_severidade,
    desc_severidade,
    cinto_seguranca,
    COUNT(*) AS total
FROM si_env_2024
WHERE (condutor = 'S' OR passageiro = 'S')
GROUP BY
    cod_severidade,
    desc_severidade,
    cinto_seguranca
ORDER BY
    cod_severidade,
    cinto_seguranca;
```

- Tabela:

sexo_condutor	cod_severidade	desc_severidade	quantidade_casos
F	0	NAO INFORMADO	6
F	1	NAO FATAL	409
F	2	FATAL	5
F	3	SEM FERIMENTOS	148
M	0	NAO INFORMADO	96
M	1	NAO FATAL	1,974
M	2	FATAL	46
M	3	SEM FERIMENTOS	993

- Gráfico:



Relação entre sexo do condutor e a severidade da condição dele mesmo em acidentes envolvendo apenas 1 condutor:

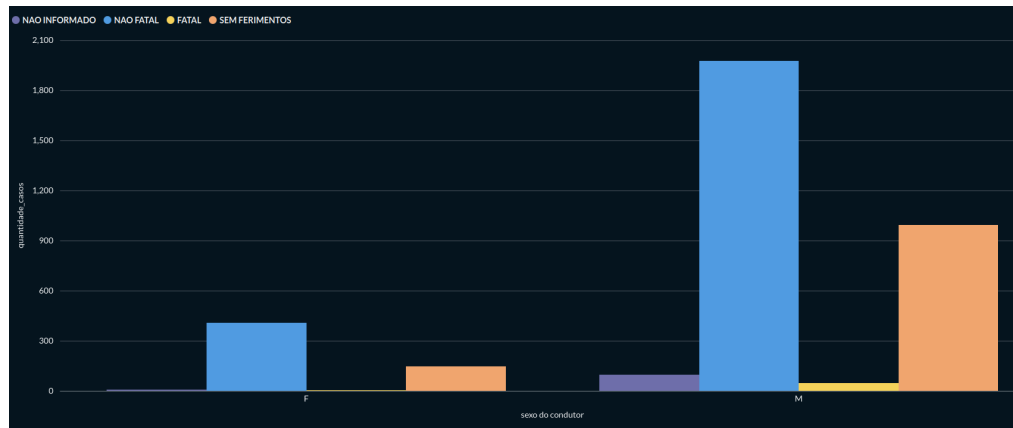
- Código SQL da consulta:

```
WITH condutores_boletim AS (
    SELECT
        No_boletim,
        SUM(CASE WHEN sexo = 'M' THEN 1 ELSE 0 END) AS qtd_homens,
        SUM(CASE WHEN sexo = 'F' THEN 1 ELSE 0 END) AS qtd_mulheres,
        COUNT(*) AS total_condutores
    FROM si_env_2024
    WHERE condutor = 'S' AND sexo IN ('M','F')
    GROUP BY No_boletim
    HAVING COUNT(*) = 2 -- exatamente 2 condutores
)
SELECT
    CASE
        WHEN qtd_homens = 2 THEN 'Homem x Homem'
        WHEN qtd_mulheres = 2 THEN 'Mulher x Mulher'
        WHEN qtd_homens = 1 AND qtd_mulheres = 1 THEN 'Homem x Mulher'
    END AS classificacao,
    COUNT(*) AS total_boletins
FROM condutores_boletim
GROUP BY classificacao
ORDER BY classificacao;
```

- Tabela:

sexo_condutor	cod_severidade	desc_severidade	quantidade_casos
F	0	NAO INFORMADO	6
F	1	NAO FATAL	409
F	2	FATAL	5
F	3	SEM FERIMENTOS	148
M	0	NAO INFORMADO	96
M	1	NAO FATAL	1,974
M	2	FATAL	46
M	3	SEM FERIMENTOS	993

- Gráfico:



Relação entre sexo do condutor e a severidade da condição dos pedestres em acidentes envolvendo apenas 1 condutor:

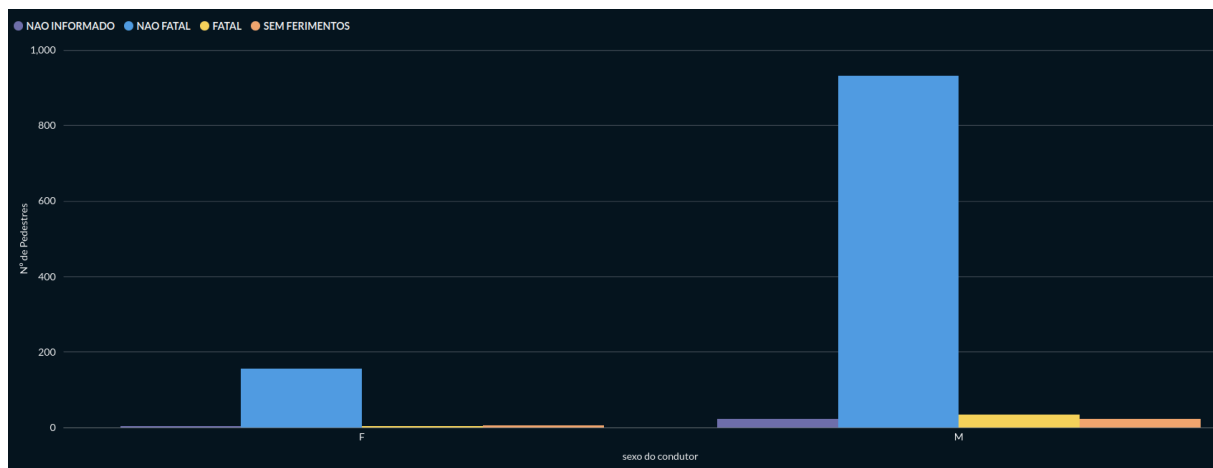
- Código SQL da consulta:

```
WITH boletins_um_condutor AS (  
    SELECT  
        No_boletim,  
        MAX(CASE WHEN condutor = 'S' AND sexo IN ('M','F') THEN sexo  
END) AS sexo_condutor  
    FROM si_env_2024  
    GROUP BY No_boletim  
    HAVING  
        SUM(CASE WHEN condutor = 'S' AND sexo IN ('M','F') THEN 1 ELSE  
0 END) = 1  
)  
SELECT  
    b.sexo_condutor,  
    e.cod_severidade,  
    e.desc_severidade,  
    COUNT(*) AS total_pedestres  
FROM boletins_um_condutor b  
JOIN si_env_2024 e  
    ON b.No_boletim = e.No_boletim  
WHERE e.pedestre = 'S'  
GROUP BY  
    b.sexo_condutor,  
    e.cod_severidade,  
    e.desc_severidade  
ORDER BY  
    b.sexo_condutor,  
    e.cod_severidade;
```

- Tabela:

sexo_condutor	cod_severidade	desc_severidade	total_pedestres
F	0	NAO INFORMADO	4
F	1	NAO FATAL	156
F	2	FATAL	4
F	3	SEM FERIMENTOS	5
M	0	NAO INFORMADO	23
M	1	NAO FATAL	932
M	2	FATAL	34
M	3	SEM FERIMENTOS	23

- Gráfico:



Relação entre sexo do condutor e ocorrência de acidentes envolvendo apenas 1 condutor:

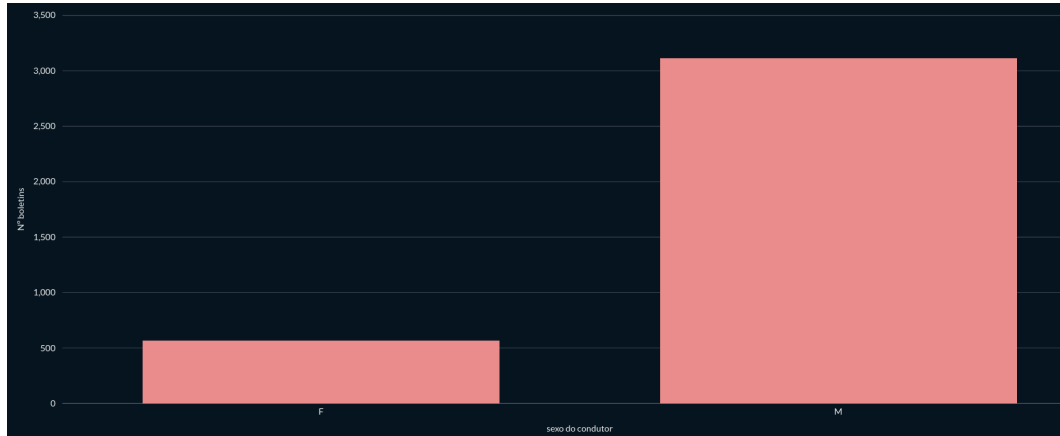
- Código SQL da consulta:

```
WITH boletins_um_condutor AS (
    SELECT
        No_boletim,
        MAX(CASE WHEN condutor = 'S' AND sexo IN ('M','F') THEN sexo
        END) AS sexo_condutor
    FROM si_env_2024
    GROUP BY No_boletim
    HAVING
        SUM(CASE WHEN condutor = 'S' AND sexo IN ('M','F') THEN 1 ELSE
        0 END) = 1
)
SELECT
    b.sexo_condutor,
    COUNT(*) AS total_boletins
FROM boletins_um_condutor b
GROUP BY
    b.sexo_condutor
ORDER BY
    b.sexo_condutor;
```

- Tabela:

sexo_condutor	total_boletins
F	568
M	3,109

- Gráfico:



Conclusões

É possível inferir diversas informações a partir da coleta e análise dos dados obtidos. Dentre essas observações, gostaria de destacar:

- A tendência predominante do envolvimento masculino nos acidentes. Na maior parte dos casos, o condutor do veículo envolvido no acidente é homem, o que indica uma tendência desse público a estar mais presente em situações como essa;
- Acidentes envolvendo 1 ou 2 veículos são os mais comuns, indicando que, em geral, acidentes de menor complexidade ocorrem com uma frequência muito maior no cotidiano em relação aos engavetamentos;
- Revela-se o quão importante é o uso do cinto de segurança, uma vez que apresenta uma fatalidade média inferior a 5% dos casos em que não há utilização do equipamento de segurança;
- A inconsistência dos dados em relação à embriaguez não permite uma análise precisa do impacto da ingestão alcoólica para as taxas de acidentes;

Links do repositório:

- <https://github.com/JoaoGabrielMacedodePaula/Trabalho-Pr-tico---ILC.git>