

Relatório – Parte 1 do Trabalho Prático

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Disciplina: Introdução a Bancos de Dados (IBD)
Semestre: 2º/2025

Integrantes: Bruno Gontijo, Pedro Pethes, Cauã Dutra

1. Fontes Utilizadas

Foram utilizados dois conjuntos de dados públicos oficiais que representam aspectos distintos, porém correlacionados, da realidade municipal brasileira:

- População municipal: obtida no Instituto Brasileiro de Geografia e Estatística (IBGE), disponível em <https://www.ibge.gov.br>
- Frota de veículos por município: obtida na Secretaria Nacional de Trânsito (SENATRAN), disponibilizada no Portal Brasileiro de Dados Abertos (<https://dados.gov.br>)

Ambos os conjuntos se referem ao ano-base de 2024 e abrangem todos os 5.571 municípios brasileiros.

2. Processo de Preparação

Os arquivos originais em formato CSV foram baixados, analisados e preparados para importação no PostgreSQL, seguindo boas práticas de limpeza e padronização de dados.

Etapas de limpeza e padronização

1. Padronização do formato: ambos os arquivos foram convertidos para o delimitador ';' e codificação UTF-8.
2. Tabela de população (pop_municipio_2024_limpo.csv):
 - Removidos pontos separadores de milhar e colunas vazias.
 - Criado o campo codigo_municipio_ibge combinando COD. UF (2 dígitos) +

COD. MUNIC (5 dígitos), totalizando 7 dígitos únicos para cada município.

- Todos os campos numéricos foram convertidos para tipo inteiro.

3. Tabela de frota (frota_municipio_2024_ok.csv):

- Padronizados os nomes dos municípios (remoção de acentos, apóstrofos e espaços duplicados).

- Corrigidos tipos de dados para valores inteiros em todas as colunas quantitativas.

- Eliminadas eventuais inconsistências de formatação e confirmada a ausência de duplicatas.

4. Integração das tabelas:

- A integração inicial foi feita por UF + nome de município normalizado, resultando em 99,21% de correspondência direta (5.527 de 5.571 municípios).

- Foram identificados 44 municípios com divergências de grafia (ex.: Parati → Paraty, Pingo D'Água → Pingo D'Água), corrigidos manualmente.

- Após essa correção, o campo codigo_municipio_ibge foi propagado para a tabela de frota, permitindo junção padronizada e consistente entre as duas fontes.

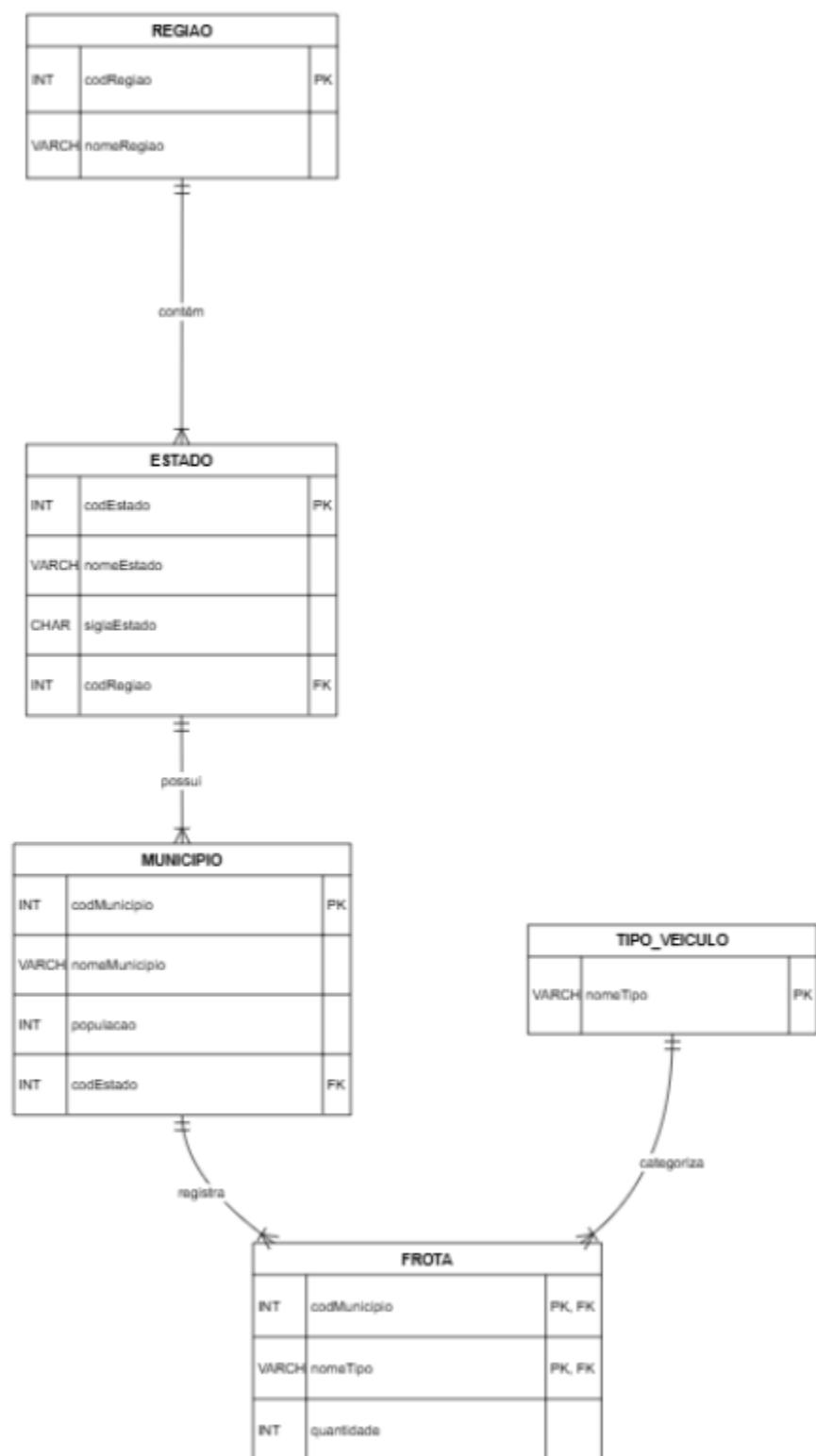
Carga no banco de dados

Os arquivos foram importados no PostgreSQL utilizando os comandos:

```
\copy populacao_municipal_2024 FROM 'pop_municipio_2024_limpo.csv'  
DELIMITER ';' CSV HEADER;  
\copy frota_municipal_2024 FROM 'frota_municipio_2024_ok.csv' DELIMITER ','  
CSV HEADER;
```

As tabelas populacao_municipal_2024 e frota_municipal_2024 estão agora integradas por meio da chave codigo_municipio_ibge.

3. Esquema Conceitual (ER/UML)



4. Dicionário de Dados

Arquivo anexado: "Dicionario_de_Dados_IBD_Parte1"

5. Conclusão

A integração entre os conjuntos de dados de população municipal e frota de veículos permitiu a construção de uma base de dados unificada, coerente e de abrangência nacional, representando informações demográficas e de mobilidade referentes ao ano de 2024. O processo de padronização possibilitou a correspondência entre as duas fontes por meio do campo `codigo_municipio_ibge`, garantindo a consistência dos registros e a unicidade de identificação dos municípios brasileiros.

Os procedimentos de limpeza eliminaram inconsistências de formatação, acentuação e codificação, assegurando a integridade e a qualidade das informações. A correspondência entre os conjuntos atingiu 99,2% de compatibilidade direta, e os casos restantes foram devidamente ajustados. O resultado é uma estrutura relacional sólida, adequada para consultas e cruzamentos de dados que associam a dimensão populacional à infraestrutura veicular de cada município.

O modelo construído viabiliza a realização de análises exploratórias e inferenciais, como a relação entre densidade populacional e tamanho da frota, distribuição regional de veículos per capita e identificação de padrões socioeconômicos a partir dos dados integrados. Além disso, o uso do PostgreSQL como Sistema Gerenciador de Banco de Dados (SGBD) garante robustez, escalabilidade e flexibilidade para futuras expansões da base ou inclusão de séries históricas.