

Relatório – Parte 2 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. Análise Exploratória dos Dados

1.1 Estatísticas Descritivas Gerais

População Municipal (2024)

A base de dados abrange todos os **5.571 municípios brasileiros**, com as seguintes características demográficas:

- **População Total Brasil:** 212.583.750 habitantes
- **População Mínima:** 854 habitantes
- **População Máxima:** 11.895.578 habitantes (São Paulo)
- **População Média:** 38.159 habitantes por município
- **População Mediana:** 11.354 habitantes

Frota de Veículos (2024)

A frota nacional registrada apresenta as seguintes características:

- **Frota Total Brasil:** 121.836.610 veículos
- **Frota Mínima:** 1 veículo
- **Frota Máxima:** 9.688.975 veículos (São Paulo)
- **Frota Média:** 21.870 veículos por município
- **Frota Mediana:** 4.671 veículos

1.2 Composição da Frota Nacional

A análise da composição da frota brasileira revela a seguinte distribuição por tipo de veículo:

Tipo de Veículo	Quantidade	Percentual
Automóveis	62.586.804	51,37%
Motocicletas	27.712.300	22,75%
Caminhonetas	9.793.689	8,04%
Caminhões	3.124.937	2,57%
Ônibus	719.418	0,59%
Outros	17.899.462	14,68%

Os automóveis representam mais da metade da frota nacional (51,37%), seguidos pelas motocicletas (22,75%). A categoria "outros" inclui veículos utilitários, tratores, reboques, semi-reboques, triciclos, quadriciclos e demais categorias especiais.

1.3 Distribuição Geográfica

Top 10 Municípios por População

Município UF População 2024

São Paulo SP 11.895.578

Rio de Janeiro RJ 6.729.894

Brasília DF 2.982.818

Fortaleza CE 2.574.412

Salvador BA 2.568.928

Belo Horizonte MG 2.416.339

Manaus AM 2.279.686

Curitiba PR 1.829.225

Recife PE 1.587.707

Goiânia GO 1.494.599

Top 10 Municípios por Frota

Município UF Frota Total

São Paulo SP 9.688.975

Rio de Janeiro RJ 3.184.115

Município UF Frota Total

Belo Horizonte MG 2.661.229

Brasília DF 2.125.733

Curitiba PR 1.757.137

Goiânia GO 1.356.770

Fortaleza CE 1.258.670

Salvador BA 1.057.055

Campinas SP 981.617

Manaus AM 927.883

1.4 Análise por Unidade Federativa

A tabela abaixo apresenta indicadores agregados por estado, ordenados por população total:

UF	Municípios	Pop. Total	Pop. Média	Frota Total	Frota Média	Veíc/Capita
SP	645	45.973.194	71.276	33.712.865	52.512	0,733
MG	853	21.322.691	24.997	13.675.774	16.108	0,641
RJ	92	17.219.679	187.170	7.842.622	87.140	0,455
BA	417	14.850.513	35.613	5.262.246	12.680	0,354
PR	399	11.824.665	29.636	8.987.629	22.753	0,760
RS	497	11.229.915	22.595	8.105.734	16.375	0,722
PE	185	9.539.029	51.562	3.646.138	20.034	0,382
CE	184	9.233.656	50.183	3.833.158	20.832	0,415
PA	144	8.664.306	60.169	2.689.204	18.938	0,310
SC	295	8.058.441	27.317	6.227.246	21.548	0,773

1.5 Padrões por Faixa Populacional

A análise por faixa populacional revela padrões consistentes:

Faixa Populacional	Nº Municípios	Veíc/Capita Médio	% Automóveis	% Motos
Até 5 mil	1.274	0,555	43,81%	29,09%
5-20 mil	2.519	0,479	39,06%	34,64%
20-50 mil	1.064	0,490	37,29%	35,40%
50-100 mil	335	0,557	40,59%	31,61%
Mais de 100 mil	335	0,598	49,58%	24,25%

Insights:

- Municípios **maiores (>100 mil hab)** têm maior taxa de motorização (0,598) e maior percentual de automóveis (49,58%)
- Municípios **médios (5-50 mil hab)** apresentam as maiores proporções de motocicletas (34-35%)
- Municípios **muito pequenos (<5 mil hab)** têm perfil similar aos grandes, possivelmente por serem centros regionais ou balneários

1.6 Análise Regional: Motocicletas vs Automóveis

A preferência por tipo de veículo varia significativamente entre regiões:

Região	Total Automóveis	Total Motos	Razão Motos/Carros	% Motos
Norte	2.124.394	2.640.047	1,243	55,41%
Nordeste	8.115.943	8.274.231	1,020	50,48%
Centro-Oeste	5.369.274	2.503.921	0,466	31,80%
Sudeste	33.488.918	10.535.075	0,315	23,93%
Sul	13.199.326	3.609.474	0,273	21,47%

Interpretação:

- Regiões **Norte e Nordeste** têm predominância de motocicletas sobre automóveis (razão > 1,0)
- No Norte, há **1,24 motos para cada carro**, representando 55,41% da frota
- Na região **Sul**, as motos representam apenas 21,47% da frota

1.7 Identificação de Outliers

Municípios com Frota Superior à População

Foram identificados **24 municípios** onde a frota de veículos supera o número de habitantes:

Top 5 Outliers Superiores:

Município	UF	População	Frota	Veíc/Capita
Ermo	SC	2.349	4.024	1,713
Santa Rita do Itueto	MG	5.973	9.083	1,521
Bom Jesus do Norte	ES	10.764	15.269	1,419
Rio Preto	MG	5.240	7.138	1,362
Chuí	RS	6.405	8.084	1,262

Municípios com Baixíssima Taxa de Motorização

Foram identificados **131 municípios** com taxa inferior a 0,1 veículos por habitante, concentrados principalmente nas regiões Norte (AM, PA, AP, RR, AC) e Nordeste (MA):

Top 5 Outliers Inferiores:

Município	UF	População	Frota	Veíc/Capita
Afuá	PA	40.246	12	0,000
Ipixuna	AM	25.458	59	0,002
Chaves	PA	21.487	34	0,002
Uiramutã	RR	15.571	109	0,007
Bagre	PA	34.633	422	0,012

2. Análise Crítica das Fontes de Dados

2.1 Qualidade e Completude: A análise exploratória via SQL revelou que as fontes possuem alta qualidade para integração, com as seguintes características estruturais:

- **Cobertura:** Ambas as fontes (IBGE e SENATRAN) cobrem 100% dos 5.571 municípios brasileiros.
- **Atualidade:** Os dados referem-se ao ano base de 2024, permitindo uma análise coerente.

2.2 Problemas de Integração e Tratamento de Dados (ETL) A principal dificuldade técnica encontrada foi a inconsistência nas chaves de identificação (nomes de municípios) entre as bases, exigindo etapas de limpeza:

- **Divergências de Grafia:** Identificamos 44 municípios (0,79% da base) com grafias divergentes nas chaves primárias textuais.
 - *Problema:* Diferenças de acentuação ("Acrelandia" vs "Acrelândia") e uso de hífen/apóstrofo ("D'Aliança").
 - *Solução:* Aplicação de funções de normalização de strings (remoção de acentos e *upper case*) e criação de uma tabela de-para manual para os casos persistentes (ex: "Parati" vs "Paraty").
- **Resultado da Integração:** Após o tratamento, foi possível realizar o **JOIN** entre as tabelas com 100% de sucesso, sem perda de registros.

2.3 Limitações Técnicas dos Dados Apesar da integridade estrutural, os dados apresentam limitações intrínsecas à metodologia de coleta das fontes originais:

- **Granularidade Temporal:** A base populacional é uma estimativa anual, enquanto a frota é um registro acumulado mensal. Isso pode gerar pequenas distorções em taxas *per capita*.
- **Local de Registro vs. Circulação:** O banco de dados reflete o domicílio fiscal do veículo, que não necessariamente corresponde ao local de circulação física. Isso explica a existência de *outliers* estatísticos (como municípios com frota superior à população) sem que isso configure erro no banco de dados.

3. Análise Integrada e Correlações

3.1 Correlação: População vs. Tamanho da Frota A integração entre as tabelas permitiu verificar uma **correlação linear positiva forte** entre o número de habitantes e o total de veículos. No entanto, ao analisar a taxa de motorização (Veículos/Habitante), observamos comportamentos distintos por faixa populacional:

- **Grandes Centros Urbanos:** Tendem a possuir taxas de motorização mais elevadas (0,6 a 0,8 veíc/hab), indicando uma concentração de frota superior à concentração populacional.
- **Municípios Médios:** Apresentam taxas intermediárias (0,4 a 0,6 veíc/hab).
- **Média Nacional:** A taxa média consolidada é de **0,57 veículos por habitante**.

3.2 Padrões de Composição da Frota (Automóveis vs. Motocicletas) A análise cruzada por Unidade Federativa (UF) e Região revelou uma inversão clara na predominância do tipo de veículo, sugerindo padrões distintos de mobilidade registrados no banco de dados:

- **Padrão Sul/Sudeste:** Predominância absoluta de **Automóveis** (>50% da frota) e baixa representatividade de motocicletas (razão Motos/Carros < 0,4).
- **Padrão Norte/Nordeste:** Predominância de **Motocicletas** sobre automóveis. Nestas regiões, a razão Motos/Carros supera 1,0, chegando a 1,24 na região Norte (55,41% da frota total).
- **Padrão Centro-Oeste:** Características híbridas com forte presença de veículos utilitários (caminhonetas), correlacionando-se com o perfil econômico da região (agropecuária) inferido pelo contexto geográfico.

3.3 Conclusão da Análise Integrada Os dados integrados demonstram que a taxa de motorização no Brasil não é uniforme. A variável "Região Geográfica" atua como um discriminante forte para prever tanto a densidade da frota (veículos per capita) quanto sua composição (motos vs. carros).

4. Conclusões

4.1 Principais Descobertas

- **Desigualdade Regional Acentuada:**
 - A razão motos/carros varia drasticamente: de 0,273 (Sul) a 1,243 (Norte).
 - Norte e Nordeste possuem frota majoritariamente de duas rodas, enquanto o Sul/Sudeste prioriza automóveis.
- **Concentração Urbana:**
 - As 10 maiores cidades concentram ~25% da frota nacional.
 - Municípios grandes apresentam taxas de motorização mais altas (0,6-0,8 veíc/hab) comparados aos médios.
- **Outliers Significativos:**
 - 24 municípios com frota > população (indicando distorção entre domicílio fiscal e residência).
 - 131 municípios com taxa < 0,1 (indicando isolamento geográfico ou sub-registro).
- **Qualidade dos Dados:**
 - 99,2% de integração automática bem-sucedida.
 - Dados consistentes para análises macro, após tratamento de strings.

4.2 Limitações do Estudo

- Análise baseada em registros formais (não captura veículos irregulares).
- Dados de população são estimativas (não censo completo).
- Não considera mobilidade intermodal (transporte público, fluvial, etc).

5. Repositório e Acesso aos Dados

Os dados organizados, scripts SQL utilizados, dicionário de dados e este relatório completo estão disponíveis no repositório:

GitHub: [brunogontij05/tp-ibd-dados-publico](https://github.com/brunogontij05/tp-ibd-dados-publico)

Conteúdo do repositório:

- /data/ - Arquivos CSV originais e processados
- /sql/ - Scripts SQL de criação de tabelas e análises
- /docs/ - Relatórios (Parte 1 e Parte 2), dicionário de dados, esquema ER
- README.md - Instruções de uso e reprodução das análises

Licença: Os dados são públicos (IBGE e SENATRAN). As análises e código desenvolvidos estão sob licença MIT.
