

Análise Integrada de Dados Públicos

População e Frota Veicular no Brasil

Um estudo comparativo e analítico abrangendo as partes 1 e 2 do projeto
da disciplina.

EQUIPE DE ALUNOS

-  Bruno Gontijo
-  Pedro Pethes
-  Cauã Dutra



Objetivo e Fontes de Dados

Objetivo Principal

Construir um banco de dados relacional robusto para análise demográfica e veicular, aplicando princípios rigorosos de normalização.

✓ Integração de Dados

Unificação de bases heterogêneas (IBGE e SENATRAN) em um único esquema.

✓ Normalização (3FN)

Implementação de modelo relacional até a 3^a Forma Normal para integridade dos dados.

✓ Análise de Correlação

Estudo comparativo entre densidade populacional e volume de frota veicular.

FONTES GOVERNAMENTAIS

IBGE



ESTIMATIVAS DE POPULAÇÃO

Referência: 2024

Municípios

5.571

Habitantes

212,5 mi

SENATRAN



FROTA DE VEÍCULOS

Referência: Julho/2024

Abrangência

Nacional

Veículos

121,8 mi

STACK TECNOLÓGICA



PostgreSQL
SGBD RELACIONAL

* pgAdmin 4

</> SQL

Esquema Relacional Normalizado



Desafios de ETL: Limpeza e Normalização

⚠ Problemas Identificados

Inconsistência em Arquivos de Origem

- ✗ **População:** Cabeçalhos "sujos" e números formatados como texto (ex: "12.345").
- ✗ **Frota:** Estrutura pivotada (veículos em colunas) dificultando consultas agregadas.

Conflitos de Integridade Referencial

A união das bases revelou divergências na grafia de municípios:

- 📍 44 municípios (0,79% da base) apresentaram divergências críticas (acentuação, hífens, apóstrofos).
- ⚠ Exemplos reais: "Parati" vs "Paraty", "Mogi Mirim" vs "Mogi-Mirim".

✓ Soluções Implementadas

• Normalização Estrutural (Unpivot)

Transformação da tabela Frota de colunas para linhas, permitindo análises mais flexíveis por tipo de veículo sem alterar o schema.

• Tratamento de Strings (PostgreSQL)

Utilização da extensão `unaccent` para ignorar acentos durante o join e criação de tabela "De-Para" manual para exceções ortográficas.

RESULTADO FINAL

100% de correspondência

A limpeza de dados representou aproximadamente 90% do esforço técnico nesta etapa.

Estratégia de Consultas SQL

Aplicação de filtros, junções, agregações e subconsultas para extração de inteligência.

1. FILTRAGEM BÁSICA

2 Consultas

- ▼ **Seleção Condicional:** Municípios filtrados por corte populacional mínimo.
- ↓ **Ordenação:** Estados rankeados por índice de motorização.

3. AGREGAÇÃO (GROUP BY)

3 Consultas

- ⌚ **Composição da Frota:** Segmentação por tipo (Automóveis 51%, Motos 22%).
- ⌚ **Totais Regionais:** Somatórios consolidados por 5 regiões e 27 UFs.

2. INTEGRAÇÃO (JOINS)

3 Consultas

- 📌 **Múltiplas Tabelas:** Junção de MUNICÍPIO ⇔ ESTADO ⇔ FROTA.
- ⇄ **Análise Cruzada:** Correlação direta entre volume populacional e frota veicular.
- gMaps **Visão Geográfica:** Distribuição consolidada por macrorregiões.

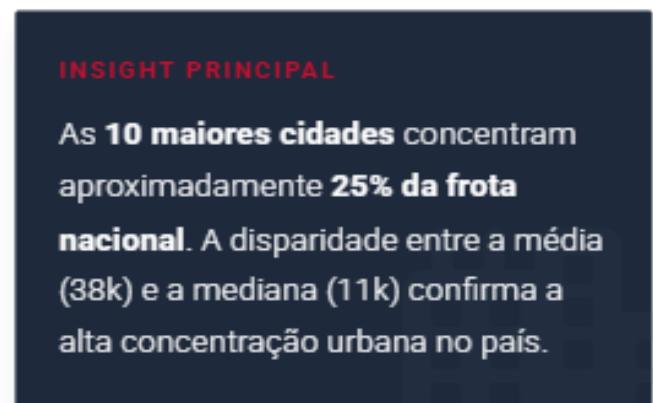
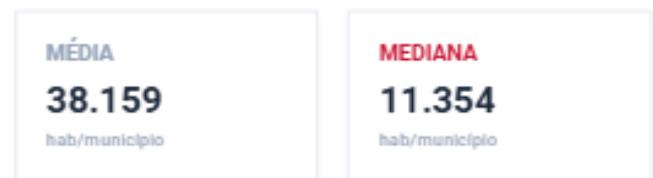
4. CONSULTAS ANINHADAS

2 Consultas

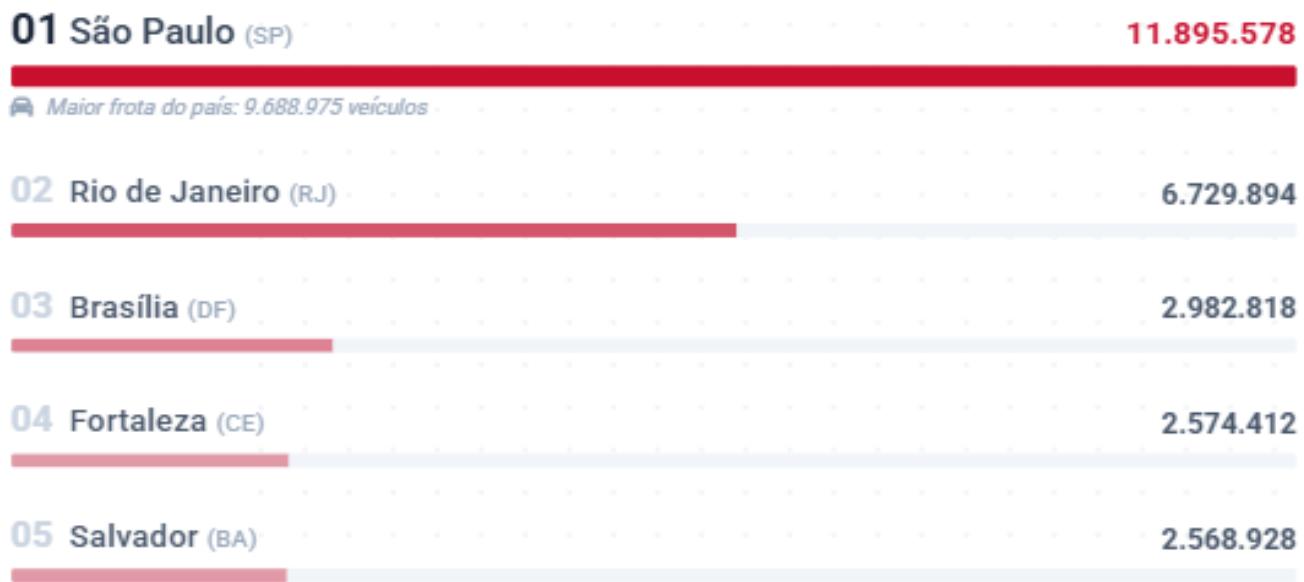
- ⌚ **Benchmarking:** Comparação dinâmica com a média nacional (0,57 v/hab).
- ⚠ **Outliers:** Identificação de 24 municípios com mais veículos que habitantes.

Concentração Urbana e Assimetria

A análise dos 5.571 municípios revela uma forte concentração populacional em poucos centros urbanos, contrastando com a maioria das cidades de pequeno porte.



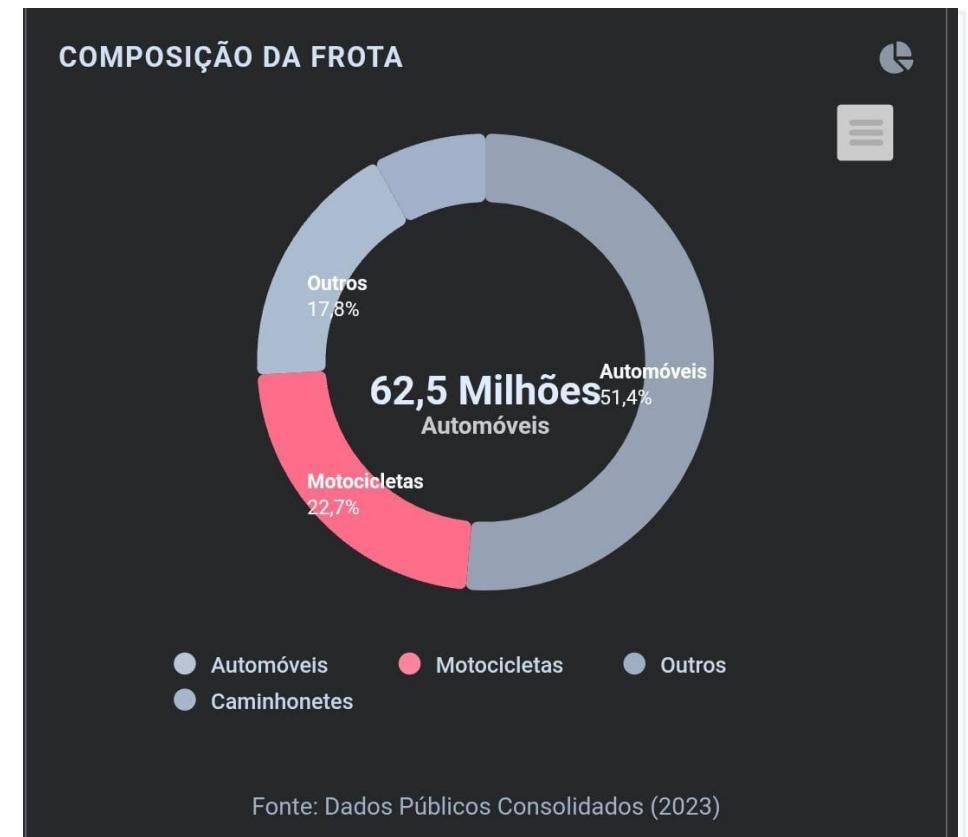
- Top 5 Maiores Municípios (População)



💡 Observa-se um degrau significativo entre São Paulo/Rio de Janeiro e as demais capitais regionais.

Correlação: População vs. Frota Veicular

"A integração de dados revela disparidades regionais drásticas na composição da frota e motorização."



Aspectos culturais e econômicos

Percebemos que a relação de frota de veículos e do número de habitantes de uma região vai além de uma relação proporcional. As tabelas mostram dados que podem refletir a cultura e a situação econômica do local

Veículos per capita - Top 10 municípios

	nome text	uf character (2)	populacao_2024 integer	frota_total integer	veiculos_per_capita numeric
1	Ermo	SC	2349	4024	1.713
2	Santa Rita do Itueto	MG	5973	9083	1.521
3	Bom Jesus do Norte	ES	10764	15269	1.419
4	Rio Preto	MG	5240	7138	1.362
5	Santa Bárbara do Monte Ver...	MG	3199	4180	1.307
6	Passa Vinte	MG	2301	3005	1.306
7	Chuí	RS	6405	8084	1.262
8	Aceguá	RS	4253	4877	1.147
9	Santa Rosa de Lima	SC	2128	2441	1.147
10	Rio Fortuna	SC	5006	5530	1.105

Explicações possíveis:

- Municípios turísticos ou balneários (segunda residência)
- Centros logísticos ou comerciais regionais
- Registro de veículos de empresas de transporte
- Frota de veículos utilitários agrícolas ou comerciais
- Belo Horizonte também aparece nesta categoria (1,101), possivelmente por ser polo econômico regional

Municípios com menor taxa de motorização

	nome text	uf character (2)	populacao_2024 integer	frota_total integer	veiculos_per_capita numeric
1	Afuá	PA	40246	12	0.000
2	Ipixuna	AM	25458	59	0.002
3	Chaves	PA	21487	34	0.002
4	Uiramutã	RR	15571	109	0.007
5	Bagre	PA	34633	422	0.012
6	Melgaço	PA	29846	372	0.012
7	Aveiro	PA	19223	365	0.019
8	Limoeiro do Aju...	PA	31778	606	0.019
9	Anajás	PA	30003	590	0.020
10	Curralinho	PA	36451	730	0.020

Explicações possíveis:

- Municípios ribeirinhos ou isolados (transporte fluvial predominante)
- Baixo desenvolvimento econômico
- Infraestrutura viária precária ou inexistente
- Possíveis problemas de registro/atualização dos dados
- Mobilidade baseada em embarcações (Afuá-PA, município insular)

	uf character (2) 	num_municipios bigint 	pop_total bigint 	pop_media numeric 	frota_total bigint 	frota_media numeric 	veiculos_per_capita_uf numeric 
1	SP	645	45973194	71276	33712865	52512	0.733
2	MG	853	21322691	24997	13675774	16108	0.641
3	RJ	92	17219679	187170	7842622	87140	0.455
4	BA	417	14850513	35613	5262246	12680	0.354
5	PR	399	11824665	29636	8987629	22753	0.760
6	RS	497	11229915	22595	8105734	16375	0.722
7	PE	185	9539029	51562	3646138	20034	0.382
8	CE	184	9233656	50183	3833158	20832	0.415
9	PA	144	8664306	60169	2689204	18938	0.310
10	SC	295	8058441	27317	6227246	21548	0.773

Conclusões e Lições Aprendidas

A análise revelou padrões demográficos distintos e validou a eficácia da modelagem relacional.

Principais Descobertas

Desigualdade Regional Acentuada

Norte/Nordeste:	Preferência por Motos (Razão > 1.2)
Sul/Sudeste:	Domínio de Carros (Razão < 0.3)

Concentração Urbana Extrema

Apenas 10 cidades concentram ~25% da frota nacional.

ℹ️ Taxa de motorização em grandes centros: 0,6-0,8 veíc./hab.

Qualidade da Integração

99.2%

MATCH AUTOMÁTICO

100%

PÓS-TRATAMENTO

Lições Técnicas

Integridade via Normalização

A aplicação da 3ª Forma Normal foi crucial para evitar redundâncias e garantir a consistência dos dados cruzados.

ETL é 90% do Trabalho

A limpeza, padronização e carga dos dados consumiram a maior parte do tempo, confirmando a realidade da engenharia de dados.

Potencial do SQL

Consultas agregadas permitiram descobrir insights (como a correlação frota/população) invisíveis nos arquivos CSV isolados.

REPOSITÓRIO DO PROJETO

🔗 github.com/brunogontij05/ibd-tp-dados-publicos

Obrigado!

Aberto a perguntas