
Pipeline de Dados e Análise Exploratória para Conformidade Ambiental Veicular: Uma Abordagem Baseada no PROCONVE L8

Felipe Leme Dias



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Monte Carmelo - MG
2026

Felipe Leme Dias

**Pipeline de Dados e Análise Exploratória para
Conformidade Ambiental Veicular: Uma
Abordagem Baseada no PROCONVE L8**

Trabalho de Conclusão de Curso apresentado
à Faculdade de Computação da Universidade
Federal de Uberlândia, Minas Gerais, como
parte dos requisitos para à obtenção do grau
de Bacharel em Sistemas de Informação.

Área de concentração: Sistemas de Informação

Orientador: Igor da Penha Natal

Monte Carmelo - MG

2026

Felipe Leme Dias

Pipeline de Dados e Análise Exploratória para Conformidade Ambiental Veicular: Uma Abordagem Baseada no PROCONVE L8

Trabalho de Conclusão de Curso apresentado à Faculdade de Computação da
Universidade Federal de Uberlândia, Minas Gerais, como parte dos requisitos para à
obtenção do grau de Bacharel em Sistemas de Informação.

Monte Carmelo, 30 de Julho de 2026.

Banca Examinadora:

Prof. Dr. Igor da Penha Natal (Orientador)
Universidade Federal de Uberlândia

Prof. Dr. Carlos Cesar Mansur Tuma
Universidade Federal de Uberlândia

Prof. Dr. Victor Sobreira
Universidade Federal de Uberlândia

À minha família, meu porto seguro em todas as horas. Aos meus avós, raiz de tudo que somos. Aos meus pais, que amaram sem medir e deram sem esperar volta. Cada um de vocês mora um pouco em cada página deste trabalho. Esta conquista é nossa, sempre foi.

Agradecimentos

Agradecer é reconhecer que nenhuma conquista é solitária. Este trabalho é também de vocês.

À minha família, e em especial aos meus avós, ao meu pai e à minha mãe: vocês tornaram possível o impossível. Abriram mão de muito para que eu não precisasse abrir mão de nada, abriram mão da minha presença em casa para que eu pudesse crescer longe dela, e nunca deixaram eu sentir o peso disso. Cada conta paga no limite, cada fim de semana que eu não voltei pra casa, cada vez que seguraram a saudade no telefone pra não me deixar preocupado, cada ligação, cada visita, cada gesto silencioso, formaram a base de tudo que construí. Me deram raiz quando eu precisava de chão, e asas quando eu precisava partir. Aprendi com vocês que amor se prova no dia a dia, na constância, no apoio que não pede palco e não se mede. Devo a vocês não só a formação acadêmica, mas a pessoa que me tornei. Esta conquista começa em vocês e a vocês volta.

Aos amigos que a graduação me deu e que a vida, espero, vai me deixar guardar. Em especial ao pessoal das kitnets de Monte Carmelo: foram vocês que tornaram a distância de casa mais suportável, que transformaram dias cansativos em noites que valeram a pena, e que fizeram de Monte Carmelo um lugar com o qual tenho muito mais do que memórias acadêmicas. Transformaram um prédio pequeno em algo muito maior: um lar. Dividimos muito mais do que experiência acadêmica: dividimos medos, risadas, madrugadas, e a certeza de que a gente ia conseguir. Levo cada um de vocês comigo.

A quem dividiu comigo não só os dias bons, mas os difíceis, que foram muitos. Você esteve presente numa fase que me formou como pessoa tanto quanto a faculdade me formou como profissional, e isso não é pouca coisa. Obrigado pelo afeto, pelo cuidado, e por tudo que construímos juntos.

Aos professores do Campus Monte Carmelo, pelo comprometimento genuíno com o ensino e pela generosidade com que compartilharam não só conhecimento, mas orientação. Em especial ao meu orientador, Igor da Penha Natal, pela confiança depositada neste trabalho.

A tudo que aconteceu fora das salas de aula, as conversas que não estavam no currículo,

as experiências que não cabem em currículo nenhum, os erros que ensinaram mais do que os acertos. A graduação foi o cenário; a vida real foi a aula.

Às equipes de Controladoria e de Tecnologia da empresa onde hoje trabalho: aprendi com vocês o que nenhum livro ensina, como é trabalhar rodeado de pessoas que elevam o seu nível todos os dias. O apoio de vocês durante este processo final fez toda a diferença.

A todos: muito obrigado.

“In God we trust, all others must bring data.”
(W. Edwards Deming)

Resumo

Este trabalho apresenta um pipeline de dados para análise da conformidade dos veículos leves comercializados no Brasil em 2026 com os limites de emissão do PROCONVE Fase L8, vigente desde 1º de janeiro de 2025. A partir da extração automatizada dos dados da tabela PBEV 2026 (INMETRO) em formato PDF, utilizando a biblioteca Python `pdfplumber`, construiu-se um *dataset* estruturado com 751 modelos/versões de 39 marcas e 17 variáveis. A Análise Exploratória de Dados (AED), conduzida em 14 seções temáticas, aplicou os limites do L8 segmentados por tipo de combustível: 700 mg/km de CO e 60 mg/km de NMOG+NOx para gasolina/flex; 500 mg/km de CO e 460 mg/km de NMOG+NOx para diesel. Os resultados indicam que, entre veículos a gasolina e flex, seis modelos/versões violam os limites vigentes (1,2% do grupo): cinco versões do CHEVROLET SPIN 1.8L, sendo quatro pela emissão de CO (800 mg/km, +14,3%) e uma pelo NMOG+NOx (62 mg/km, +3,3%), e o GWM TANK 300 PHEV, que excede o NMOG+NOx com 73 mg/km (+21,7%). A análise das correlações entre variáveis revela que CO e NMOG+NOx têm correlação quase nula ($r = 0,11$), indicando que sistemas de pós-tratamento modernos os controlam de forma tecnicamente desacoplada. Demonstra-se ainda que eficiência energética (PBE) e conformidade de poluentes (L8) são métricas independentes, e que veículos plug-in e elétricos apresentam emissões fósseis significativamente inferiores à combustão convencional. O *dataset* processado e o dicionário de variáveis são disponibilizados como subprodutos desta pesquisa.

Palavras-chave: PROCONVE L8, Emissões veiculares, PBEV, Análise exploratória de dados, Conformidade ambiental.

Abstract

This work analyzes the compliance of light vehicles sold in Brazil in 2026 with the emission limits of PROCONVE Phase L8, which came into effect on January 1, 2025. By automatically extracting data from the PBEV 2026 (INMETRO) table in PDF format using the Python library `pdfplumber`, a structured dataset was built with 751 models/versions from 39 brands and 17 variables. The Exploratory Data Analysis (EDA), conducted in 14 thematic sections, applied L8 limits segmented by fuel type: 700 mg/km CO and 60 mg/km NMOG+NO_x for gasoline/flex; 500 mg/km CO and 460 mg/km NMOG+NO_x for diesel. The results show that, among gasoline and flex vehicles, six models/versions violate the current limits (1.2% of the group): five versions of the CHEVROLET SPIN 1.8L, four due to CO emissions (800 mg/km, +14.3%) and one due to NMOG+NO_x (62 mg/km, +3.3%), and the GWM TANK 300 PHEV, which exceeds NMOG+NO_x with 73 mg/km (+21.7%). Correlation analysis reveals that CO and NMOG+NO_x are nearly uncorrelated ($r = 0.11$), indicating that modern aftertreatment systems control them in a technically decoupled manner. The analysis also shows that energy efficiency (PBE rating) and pollutant compliance (PROCONVE L8) are independent metrics, and that plug-in and electric vehicles present significantly lower fossil emissions than conventional combustion. The processed dataset and data dictionary are made available as by-products of this research.

Keywords: PROCONVE L8. Vehicle emissions. PBEV. Exploratory data analysis. Environmental compliance..

Lista de ilustrações

Figura 1 – Fluxo metodológico aplicado nesse trabalho.	28
Figura 2 – Composição da frota PBEV 2026	37
Figura 3 – Dispersão CO × NMOG+NOx com limites L8	38
Figura 4 – Distribuição de CO – Gasolina e Flex	39
Figura 5 – Distribuição de NMOG+NOx	40
Figura 6 – Top 20 maiores emissores de CO	41
Figura 7 – Conformidade por marca	42
Figura 8 – Conformidade por categoria de veículo	43
Figura 9 – CO ₂ e consumo por tecnologia	47
Figura 10 – Nota Verde INMETRO × análise própria	48
Figura 11 – Classificação PBE versus emissões e conformidade L8	49
Figura 12 – Matriz de correlação entre variáveis numéricas	50
Figura 13 – Síntese dos achados	53

Lista de tabelas

Tabela 1 – Limites de emissão PROCONVE L7 e L8	21
Tabela 2 – Comparação entre abordagens relacionadas	25
Tabela 3 – Mapeamento de colunas do PBEV 2026	30
Tabela 4 – Colunas calculadas pelo pipeline de extração	31
Tabela 5 – Schema do dataset PBEV 2026	34
Tabela 6 – Estatísticas gerais do dataset PBEV 2026	36
Tabela 7 – Composição da frota por propulsão e combustível	37
Tabela 8 – Estatísticas de CO para gasolina e flex	39
Tabela 9 – Ranking de violadores PROCONVE L8: Gasolina/Flex	42
Tabela 10 – CO ₂ fóssil e consumo por tecnologia	47
Tabela 11 – Nota Verde INMETRO × conformidade calculada	48
Tabela 12 – Correlações de Pearson: todos os pares	51

Lista de siglas

AED Análise Exploratória de Dados

CETESB Companhia Ambiental do Estado de São Paulo

CO Monóxido de Carbono

CO₂ Dióxido de Carbono

CONAMA Conselho Nacional do Meio Ambiente

GEE Gases de Efeito Estufa

HEV Veículo Híbrido Elétrico (*Hybrid Electric Vehicle*)

IARC Agência Internacional de Pesquisa em Câncer (*International Agency for Research on Cancer*)

IBAMA Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis

INMETRO Instituto Nacional de Metrologia, Qualidade e Tecnologia

IQAr Índice de Qualidade do Ar

MP Material Particulado

NMOG Hidrocarbonetos Orgânicos Não-Metânicos (*Non-Methane Organic Gas*)

NO_x Óxidos de Nitrogênio

PBE Programa Brasileiro de Etiquetagem

PBEV Programa Brasileiro de Etiquetagem Veicular

PHEV Veículo Híbrido Elétrico Plug-in (*Plug-in Hybrid Electric Vehicle*)

PROCONVE Programa de Controle da Poluição do Ar por Veículos Automotores

THC Hidrocarbonetos Totais (*Total Hydrocarbons*)

ZBE Zona de Baixas Emissões

Sumário

1	INTRODUÇÃO	16
1.1	Objetivo Geral	17
1.2	Objetivos Específicos	17
1.3	Justificativa	18
1.4	Estrutura do Trabalho	18
2	FUNDAMENTAÇÃO TEÓRICA E TRABALHOS RELACI- ONADOS	19
2.1	Poluição Atmosférica por Veículos Automotores	19
2.2	PROCONVE: Histórico e Fases	20
2.3	PROCONVE Fase L8	21
2.4	Programa Brasileiro de Etiquetagem Veicular (PBEV)	21
2.5	Classificação PBE e Nota Verde	22
2.6	Análise Exploratória de Dados	22
2.7	Extração de Dados de PDF	24
2.8	Trabalhos Relacionados	24
3	MÉTODO	27
3.1	Visão Geral do Método	27
3.2	Fonte de Dados	29
3.2.1	Considerações Sobre a Inviabilidade do Uso da FENABRAVE	29
3.3	Extração de Dados do PDF	29
3.3.1	Biblioteca utilizada	29
3.3.2	Problema do cabeçalho multi-linha	30
3.3.3	Filtro de linhas válidas	30
3.4	Tratamento e Enriquecimento dos Dados	31
3.4.1	Colunas calculadas	31
3.5	CrITÉrios de Conformidade L8 por Combustível	31

3.6	Estrutura do Dataset	32
3.6.1	Pré-processamento da fonte de dados em PDF	33
3.6.2	Dataset resultante e transformações aplicadas	33
3.7	Ferramentas e Reprodutibilidade	34
4	RESULTADOS DA ANÁLISE EXPLORATÓRIA	36
4.1	Visão Geral do Dataset	36
4.2	Composição da Frota	36
4.3	Conformidade com PROCONVE L8	37
4.4	Distribuição de CO: Gasolina e Flex	38
4.5	Distribuição de NMOG+NOx	39
4.6	Top 20 Maiores Emissores de CO	40
4.7	Análise por Marca	40
4.8	Diagnóstico de Categoria	41
4.8.1	Distribuição da frota por categoria	41
4.8.2	Categorias e perfil de emissão	42
4.8.3	Implicações regulatórias por categoria	43
4.9	Diagnóstico Ambiental	44
4.9.1	Significância dos poluentes identificados para a saúde	44
4.9.2	Contexto da conformidade no primeiro ano do L8	44
4.9.3	Impactos além do domínio ambiental	44
4.9.4	Mobilidade urbana, qualidade do ar e zonas de baixas emissões	45
4.9.5	Custos ao sistema de saúde pública	46
4.10	CO₂ Fóssil e Consumo Energético por Tecnologia	46
4.11	Nota Verde INMETRO × Conformidade Calculada	47
4.12	Classificação PBE × Conformidade L8	48
4.13	Correlações entre Variáveis	49
4.13.1	Matriz de correlação geral: todos os pares	49
4.13.2	Correlações de senso comum	49
4.13.3	Análise de causalidade	52
4.14	Síntese dos Achados	52
4.15	Análise de Cenários	54
4.15.1	Cenário 1: Aplicação de um limite de CO mais restritivo	54
4.15.2	Cenário 2: Estimativa de excedente de CO por quilometragem	54
4.15.3	Cenário 3: Conformidade total com atualização dos modelos violadores	55
4.15.4	Cenário 4: Endurecimento do limite de NMOG+NOx	55
4.15.5	Cenário 5: Impacto de deslocamento da frota para tecnologias de baixa emissão	55

5	CONCLUSÃO	57
5.1	Cumprimento dos Objetivos	57
5.2	Principais Contribuições	58
5.3	Limitações	58
5.4	Trabalhos Futuros	59
5.5	Considerações Finais	59
	REFERÊNCIAS	61

Introdução

A qualidade do ar em áreas urbanas é uma das principais preocupações de saúde pública no Brasil e no mundo. O setor de transportes, em especial os veículos automotores leves, é um dos maiores responsáveis pela emissão de poluentes atmosféricos, como o monóxido de carbono (CO), os hidrocarbonetos orgânicos não-metânicos e óxidos de nitrogênio (NMOG+NO_x) e o material particulado (MP). Esses poluentes são quimicamente reativos e possuem efeitos tóxicos diretos sobre a saúde humana: o CO inibe o transporte de oxigênio no sangue; o NO_x favorece a formação de ozônio troposférico; e os hidrocarbonetos contribuem para a névoa fotoquímica (WHO, 2021). Esses impactos diferem fundamentalmente do CO₂, que é um gás de efeito estufa (GEE) com impacto climático global, mas sem toxicidade direta em concentrações ambientais.

Para controlar esses poluentes, o Brasil instituiu em 1986 o Programa de Controle da Poluição do Ar por Veículos Automotores (PROCONVE), regulamentado pelo Conselho Nacional do Meio Ambiente (CONAMA) (CONAMA, 1986). Ao longo de quase quatro décadas, o programa estabeleceu fases progressivamente mais restritivas para veículos leves (L1 a L8), acompanhando os avanços tecnológicos da indústria automobilística. A **Fase L8**, a mais rigorosa já adotada no país, entrou em vigor em 1º de janeiro de 2025. Em relação à fase anterior (L7), os limites de CO foram reduzidos em 46% (de 1.300 para 700 mg/km, para gasolina/flex) e os de NMOG+NO_x em 25% (de 80 para 60 mg/km) (IBAMA, 2024).

Nesse contexto, este trabalho investiga a seguinte questão:

De que forma um pipeline automatizado de ingestão e tratamento de dados semiestruturados pode ser projetado para validar computacionalmente restrições regulatórias complexas e quantificar anomalias em bases de dados públicas?

Para responder a essa questão, adotou-se como objeto de estudo e validação experimental o Programa Brasileiro de Etiquetagem Veicular (PBEV), coordenado pelo INMETRO. O PBEV constitui uma base de dados pública e periódica contendo os valores certificados de emissões e consumo de todos os veículos leves homologados no país (INMETRO,

2026). Contudo, por ser disponibilizado em formato de tabela semiestruturada (PDF) e sem o suporte de APIs corporativas, apresenta desafios típicos de extração, limpeza e modelagem de dados. A edição de janeiro de 2026 (Revisão 04), que abrange 751 modelos/versões de 39 marcas distintas, foi utilizada como o dataset alvo para a homologação do pipeline desenvolvido.

Uma análise de conformidade rigorosa com o L8 exige uma distinção fundamental: os limites de emissão diferem por tipo de combustível. Para gasolina e flex, os limites são $\text{CO} \leq 700 \text{ mg/km}$ e $\text{NMOG}+\text{NOx} \leq 60 \text{ mg/km}$. Para diesel, os limites são $\text{CO} \leq 500 \text{ mg/km}$ e $\text{NMOG}+\text{NOx} \leq 460 \text{ mg/km}$. Aplicar os limites de gasolina/flex a veículos a diesel geraria centenas de falsos positivos. Este trabalho trata essa segmentação como etapa metodológica obrigatória.

1.1 Objetivo Geral

Projetar, implementar e avaliar um pipeline computacional automatizado de engenharia de dados para a extração, tratamento e validação de restrições de integridade regulatória em bases de dados semiestruturadas, utilizando como cenário experimental a detecção e quantificação programática de anomalias na frota veicular brasileira de 2026 em relação à Fase L8 do PROCONVE.

1.2 Objetivos Específicos

- ❑ Extrair automaticamente os dados de emissões da tabela PBEV 2026 em formato PDF por meio da biblioteca `pdfplumber`;
- ❑ Construir um *dataset* estruturado com 751 modelos/versões e 17 variáveis, acompanhado de um dicionário de variáveis documentado;
- ❑ Aplicar os limites do PROCONVE L8 corretamente segmentados por tipo de combustível (gasolina/flex e diesel);
- ❑ Calcular o percentual de excesso de CO e NMOG+NOx em relação aos respectivos limites por combustível;
- ❑ Conduzir uma Análise Exploratória de Dados (AED) com 14 seções temáticas, gerando 12 visualizações;
- ❑ Comparar a conformidade calculada com a Nota Verde INMETRO e a classificação PBE;

- Contextualizar os achados em dimensões complementares: implicações para a saúde pública, mobilidade urbana (zonas de baixas emissões) e cenários de evolução regulatória e tecnológica da frota.

1.3 Justificativa

A introdução de uma fase regulatória mais restritiva não garante que todos os modelos em comercialização se tornem imediatamente conformes. Fabricantes têm prazos para descontinuar ou atualizar motorizações antigas, e modelos com tecnologia defasada podem permanecer nas concessionárias por meses ou anos após a entrada em vigor da nova fase. Identificar quais modelos ainda violam os limites do L8 e quantificar o excesso é relevante para consumidores, formuladores de políticas públicas e órgãos de fiscalização.

Do ponto de vista metodológico, os dados do PBEV são publicados apenas em formato PDF, sem interface estruturada para exportação. Isso cria uma barreira técnica para análises sistemáticas de conformidade. Este trabalho supera essa barreira ao desenvolver um *pipeline* de extração automatizada e disponibilizar o *dataset* processado e documentado como subproduto da pesquisa.

1.4 Estrutura do Trabalho

O Capítulo 2 apresenta a fundamentação teórica sobre poluição veicular, o histórico do PROCONVE e a metodologia de AED, além dos trabalhos relacionados. O Capítulo 3 descreve o *pipeline* de extração e as decisões metodológicas adotadas. O Capítulo 4 apresenta os resultados da análise exploratória. O Capítulo 5 conclui o trabalho, discute limitações e aponta direções para pesquisas futuras.

Fundamentação Teórica e Trabalhos Relacionados

Este capítulo traz a base teórica do trabalho: primeiro diferencia gases de efeito estufa de poluentes locais (CO, NO_x, NMOG, MP), depois apresenta o PROCONVE e sua Fase L8 (limites de emissão vigentes desde 2025), o PBEV como fonte de dados e a diferença entre Classificação PBE e Nota Verde, seguido pelas técnicas de Análise Exploratória de Dados e extração de PDF usadas na metodologia, fechando com os trabalhos relacionados e a lacuna que esta pesquisa preenche: uma análise pública e sistemática de conformidade com o L8, modelo a modelo.

2.1 Poluição Atmosférica por Veículos Automotores

Os veículos automotores emitem dois grupos distintos de substâncias com impactos ambientais e à saúde muito diferentes. O primeiro grupo são os **gases de efeito estufa (GEE)**, principalmente o dióxido de carbono (CO₂), cujo impacto é climático e global. O segundo grupo são os **poluentes locais**, que incluem o monóxido de carbono (CO), os hidrocarbonetos orgânicos não-metânicos (NMOG), os óxidos de nitrogênio (NO_x), o material particulado (MP) e o dióxido de enxofre (SO₂). Esses poluentes têm efeitos tóxicos diretos sobre a saúde humana e provocam degradação ambiental em nível regional (WHO, 2021).

O **monóxido de carbono (CO)** é um gás inodoro e incolor resultante da combustão incompleta. Ele se liga à hemoglobina no sangue com afinidade 200 vezes maior que o oxigênio, formando carboxiemoglobina e reduzindo a capacidade de transporte de oxigênio. Em concentrações elevadas, pode causar danos neurológicos e morte. O CO é produzido principalmente por motores a gasolina em regimes ricos de mistura combustível (WHO, 2021).

Os **óxidos de nitrogênio (NO_x)**, principalmente NO e NO₂, são formados na combustão a alta temperatura. O NO_x reage com compostos orgânicos voláteis na presença

de luz solar para formar ozônio (O_3) troposférico, componente principal do *smog* fotoquímico. O NO_2 possui toxicidade pulmonar direta e é associado ao aumento de doenças respiratórias (WHO, 2021). Além disso, o NO_x contribui para a chuva ácida e para a eutrofização de ecossistemas.

Os **hidrocarbonetos não-metânicos (NMOG)** são compostos orgânicos emitidos tanto durante a combustão quanto por evaporação de combustível. Junto com o NO_x , são precursores do ozônio troposférico e do material particulado secundário. Alguns NMOG, como o benzeno e o 1,3-butadieno, são reconhecidamente carcinogênicos (WHO, 2021).

É importante distinguir o CO dos CO_2 : embora a redução de CO_2 seja fundamental para o combate às mudanças climáticas, reduzir CO_2 não implica necessariamente reduzir CO ou NO_x , e vice-versa. São problemas ambientais distintos que requerem políticas específicas. Esta distinção é relevante para a análise proposta neste trabalho, que se concentra nos poluentes locais regulados pelo PROCONVE L8.

2.2 PROCONVE: Histórico e Fases

O Programa de Controle da Poluição do Ar por Veículos Automotores (**PROCONVE**) foi criado pela Resolução CONAMA nº 18, de 6 de maio de 1986, sendo pioneiro na América Latina (CONAMA, 1986). A criação do PROCONVE foi motivada pela crescente degradação da qualidade do ar em grandes centros urbanos brasileiros, especialmente São Paulo, onde a frota veicular já causava episódios críticos de poluição.

O programa adota uma estratégia de fases progressivas, em que cada fase estabelece limites de emissão mais restritivos e tecnologias obrigatórias. Para veículos leves de passageiros (denominados Fase L), as fases evoluíram da seguinte forma (IBAMA, 2024):

- ❑ **L1 (1988)**: Primeira fase, estabeleceu limites iniciais e tornou obrigatório o catalisador de oxidação;
- ❑ **L2 (1992)**: Reduziu limites e introduziu a injeção eletrônica de combustível;
- ❑ **L3 (1997–2000)**: Catalisador de três vias e sonda lambda tornaram-se obrigatórios;
- ❑ **L4 (2003–2005)**: Aproximação com os padrões EURO 3, limites mais restritivos para NO_x e HC;
- ❑ **L5 (2009–2013)**: Equivalência ao EURO 4, com diagnóstico de bordo (OBD) obrigatório;
- ❑ **L6 (2013)**: Transição com aprimoramento das tecnologias de pós-tratamento;
- ❑ **L7 (2022)**: Equivalente ao EURO 5, reduções significativas em CO e NMOG+ NO_x ;

□ **L8 (2025):** Fase atual, a mais restritiva da história do programa, com implantação em três etapas (2025, 2027 e 2029), introdução do diagnóstico de bordo OBD Br-3, testes de emissões em condições reais de condução (RDE) e tecnologia obrigatória de captura de vapores de combustível (ORVR) para veículos a gasolina e flex.

Ao longo dessas quatro décadas, o PROCONVE contribuiu decisivamente para a melhoria da qualidade do ar nas cidades brasileiras. Estudos do CETESB demonstram que as emissões médias de CO por veículo leve reduziram-se em mais de 95% entre 1986 e 2023, mesmo com o crescimento da frota (CETESB, 2024).

2.3 PROCONVE Fase L8

A Fase L8 do PROCONVE entrou em vigor em 1º de janeiro de 2025 para veículos leves, representando a mais significativa redução de limites já imposta no país (IBAMA, 2024). A Tabela 1 apresenta os limites estabelecidos pela Fase L8 comparados com a fase anterior (L7).

Tabela 1 – Limites de emissão do PROCONVE Fases L7 e L8 para veículos leves.

Poluente	L8 Gasolina/Flex	L8 Diesel	L7 Gasolina	Redução
CO	700 mg/km	500 mg/km	1.300 mg/km	−46%
NMOG + NOx	60 mg/km	460 mg/km	80 mg/km	−25%
NOx	0,06 g/km	0,08 g/km	0,08 g/km	−25%
THC	60 mg/km	—	80 mg/km	−25%
Material Part.	4,5 mg/km	4,5 mg/km	6 mg/km	−25%

Fonte: IBAMA (IBAMA, 2024). Nota: limites de NMOG+NOx para diesel (460 mg/km) são distintos dos limites para gasolina/flex (60 mg/km).

Um aspecto crítico da Tabela 1 é a diferença substancial entre os limites de NMOG+NOx para diesel (460 mg/km) e para gasolina/flex (60 mg/km). Essa diferença existe porque os motores a diesel possuem uma arquitetura de combustão fundamentalmente diferente: operam em mistura pobre (*lean burn*) (RAPP et al., 2016), produzindo menos CO, mas mais NOx e partículas. Os limites foram calibrados de acordo com as características técnicas de cada tecnologia de propulsão (IBAMA, 2024).

Qualquer análise de conformidade que aplique os limites de gasolina/flex a veículos a diesel produzirá falsos positivos em escala. Essa é uma distinção metodológica fundamental que este trabalho aborda explicitamente.

2.4 Programa Brasileiro de Etiquetagem Veicular (PBEV)

O **Programa Brasileiro de Etiquetagem Veicular (PBEV)**, coordenado pelo INMETRO, foi criado no início dos anos 2000 como instrumento de transparência e orientação ao consumidor. O programa certifica anualmente todos os modelos de veículos leves autorizados para comercialização no Brasil, avaliando indicadores de eficiência energética e emissões (INMETRO, 2024).

A tabela anual do PBEV é publicada em formato PDF e contém, para cada modelo/versão certificada, informações como: categoria do veículo, marca, modelo, versão, motorização, tipo de propulsão, transmissão, combustível, valores de emissão de CO, NMOG+NO_x e CO₂ fóssil, consumo energético em MJ/km e a classificação de eficiência energética (PBE).

A partir de 2025, com a vigência do L8, a tabela PBEV passou a incluir também a **Nota Verde**, um indicador de conformidade ambiental atribuído pelo INMETRO durante a certificação, que classifica o veículo de A (melhor) a C (pior) com base em critérios que contemplam todos os poluentes regulados (INMETRO, 2024).

A edição PBEV 2026 (Revisão 04, janeiro de 2026) cobre 751 modelos/versões de 39 marcas, em 16 categorias de veículos, constituindo a base de dados primária deste trabalho (INMETRO, 2026).

2.5 Classificação PBE e Nota Verde

O sistema de classificação do PBEV é composto por dois indicadores distintos:

Classificação PBE: avalia a *eficiência energética* do veículo, isto é, a quantidade de energia consumida para percorrer uma quilometragem (MJ/km). A classificação vai de A (mais eficiente) a E (menos eficiente) e é definida por faixas de consumo dentro de cada categoria de veículo. Um veículo eficiente (PBE A) pode, paradoxalmente, ainda emitir poluentes locais em quantidade acima do limite L8, e vice-versa. Eficiência energética e conformidade de poluentes são métricas independentes.

Nota Verde: avalia a *conformidade ambiental* em relação aos padrões PROCONVE vigentes. A classificação também é A, B ou C, e é atribuída pelo fabricante durante o processo de certificação com base em todos os poluentes regulados (CO, NMOG+NO_x, NO_x, MP) (INMETRO, 2024). A Nota Verde utiliza critérios e ponderações internas ao processo de homologação que não são integralmente públicos. Veículos com Nota A apresentaram conformidade plena; Nota B indica conformidade parcial; Nota C indica não-conformidade ou proximidade aos limites.

2.6 Análise Exploratória de Dados

A **Análise Exploratória de Dados (AED)** é uma abordagem metodológica introduzida por Tukey (1977) que visa compreender a estrutura, distribuição e padrões presentes em um conjunto de dados antes da aplicação de modelos preditivos ou inferenciais. A AED combina estatísticas descritivas com visualizações gráficas para revelar tendências, anomalias, concentrações e relações entre variáveis.

As principais técnicas de AED utilizadas neste trabalho incluem:

- ❑ **Distribuição de frequências:** histogramas e gráficos de barras para compreender a composição da frota por propulsão, combustível e categoria;
- ❑ **Estatísticas descritivas:** média, mediana, mínimo, máximo e percentis para caracterizar as distribuições de CO e NMOG+NOx;
- ❑ **Gráficos de dispersão:** para identificar a relação entre CO e NMOG+NOx e visualizar violadores em relação aos limites L8;
- ❑ **Rankings:** ordenação dos maiores emissores de CO e identificação dos modelos violadores;
- ❑ **Análise de correlação:** matriz de correlação de Pearson entre as variáveis numéricas;
- ❑ **Comparações por grupo:** análise de conformidade segmentada por marca, categoria e tipo de combustível;
- ❑ **Diagrama de caixa (*boxplot*):** representa a distribuição de uma variável numérica por meio de cinco estatísticas (mínimo, primeiro quartil, mediana, terceiro quartil e máximo), identificando valores atípicos e sendo robusto a distribuições assimétricas;
- ❑ **Gráfico de violino (*violin plot*):** extensão do boxplot que adiciona ao contorno uma estimativa de densidade (*kernel density estimation*), revelando bimodalidade e assimetrias que o boxplot simples não captura;
- ❑ **Mapa de calor (*heatmap*):** codifica valores em uma matriz bidimensional por gradiente de cor, permitindo identificar padrões em conjuntos multivariados; na forma de mapa de calor de correlação, representa todos os pares de variáveis simultaneamente, com azul indicando correlação negativa e vermelho indicando positiva;
- ❑ **Mapa de calor anotado (*annotated heatmap*):** variante do heatmap que adiciona rótulos numéricos a cada célula, combinando a percepção visual do gradiente com a precisão de uma tabela, sendo indicado para matrizes de contingência onde a leitura exata de cada célula é necessária;

- ❑ **Painel de síntese (*summary card*)**: visualização não-estatística que consolida os principais achados em linguagem direta com codificação por cor temática, com a função de tornar os resultados acessíveis a públicos não especializados sem sacrificar a precisão.

A AED é particularmente adequada para este trabalho porque o objetivo é *descrever e diagnosticar* o estado de conformidade da frota, não construir modelos preditivos. O resultado esperado é um conjunto de visualizações e achados que respondam à pergunta de pesquisa de forma direta e reprodutível.

2.7 Extração de Dados de PDF

Os dados do PBEV são disponibilizados pelo INMETRO exclusivamente em formato PDF, sem API ou exportação tabular. Para viabilizar a análise, é necessário extrair as tabelas contidas no PDF de forma automatizada.

A biblioteca **pdfplumber** (BIGELOW, 2023) é uma ferramenta Python de código aberto que extrai conteúdo detalhado de PDFs digitalmente gerados (não escaneados). Ela identifica tabelas a partir da geometria das linhas e da posição dos caracteres na página, sem depender de OCR (*Optical Character Recognition*). Para PDFs com tabelas bem estruturadas como a do PBEV, o **pdfplumber** oferece precisão superior às alternativas baseadas em OCR (BIGELOW, 2023).

Uma complicação frequente em PDFs com tabelas é o cabeçalho multi-linha com células mescladas, que confunde extratores automáticos. A abordagem adotada neste trabalho para superar essa limitação é descrita em detalhe no Capítulo 3.

2.8 Trabalhos Relacionados

Diversos estudos têm analisado emissões veiculares no Brasil e a efetividade do PROCONVE, mas a maior parte utiliza abordagens distintas da proposta neste trabalho. A Tabela 2 sistematiza as principais diferenças entre as abordagens encontradas na literatura e a proposta aqui apresentada.

Carvalho (2011) analisa as emissões relativas de poluentes do transporte motorizado de passageiros em grandes centros urbanos brasileiros. Publicado pelo IPEA como Texto para Discussão, o estudo propõe uma metodologia de cálculo de emissões com base em matrizes de deslocamento modal e fatores de emissão unitários por tipo de veículo, abrangendo CO, NOx e material particulado. O trabalho evidencia que políticas de incentivo ao transporte coletivo são complementares ao controle de emissões por veículo (a abordagem do PROCONVE) para a melhoria da qualidade do ar metropolitana. Em relação ao presente trabalho, a diferença fundamental é metodológica: Carvalho (2011) opera na

Tabela 2 – Comparação entre abordagens relacionadas e a proposta deste trabalho.

Referência	Foco principal	Dados utilizados	Conformidade L8?
Carvalho (IPEA, 2011)	Emissões por modo de transporte em metrô-poles	Frota circulante e matrizes modais	Não
Jacondino (UFRGS, 2005)	Quantificação de emissões via simulação de tráfego	Simuladores de tráfego e fatores de emissão	Não
SIMEA (Blucher, 2017)	Eficiência energética da frota leve	PBEV (CO ₂ e MJ/km)	Não
IEMA (2025)	Inventário nacional 1980–2024	Fatores de emissão e combustível	Não
Relatórios CETESB	Qualidade do ar, tendências históricas	Estações de monitoramento	Não
TRUE/ICCT (2025)	Emissões reais em uso em São Paulo	Sensoriamento remoto (323 mil medições)	Não
Este trabalho	Conformidade modelo a modelo com L8	PBEV 2026 (poluentes locais)	Sim, por combustível

Fonte: Elaborado pelo autor.

escala da frota circulante agregada e dos modos de transporte, enquanto aqui se opera na escala do modelo/versão certificado, com dados primários do PBEV, identificando violadores individuais.

Jacondino (2005), em dissertação de mestrado defendida na Universidade Federal do Rio Grande do Sul (UFRGS), avalia o uso de simuladores de tráfego microscópicos combinados a modelos de fatores de emissão para estimar as emissões veiculares associadas a intervenções na malha viária. O estudo demonstra a viabilidade de simuladores de tráfego como ferramenta de quantificação de emissões em nível de corredor viário, porém opera com fatores de emissão médios por fase do PROCONVE vigente à época (anterior à L8) e não com dados de certificação por modelo/versão. Em relação ao presente trabalho, a diferença é dupla: a escala de análise (fluxo de tráfego agregado *versus* modelo certificado individual) e a fonte de dados (simulação *versus* dados oficiais do PBEV).

Publicado no XXV Simpósio Internacional de Engenharia Automotiva (SIMEA), o trabalho disponível nos proceedings Blucher utiliza os dados do PBEV, a mesma fonte primária deste estudo, para analisar a variação de eficiência energética da frota de veículos leves comercializados no Brasil ao longo de oito anos (LOPES et al., 2017). Os resultados apontam redução de 3% no consumo energético médio da frota (de 2,27 para 2,20 MJ/km), associada principalmente à expansão dos veículos flex e à introdução do INOVAR-AUTO. A abordagem metodológica é a mais próxima da adotada neste trabalho (extração e análise de dados tabulares do PBEV), porém com foco exclusivo nas variáveis de eficiência energética e CO₂, sem avaliação dos poluentes locais (CO e NMOG+NO_x) nem da conformidade modelo a modelo com os limites do PROCONVE.

O Inventário Nacional de Emissões Atmosféricas por Veículos Automotores Rodoviários, elaborado pelo Instituto de Energia e Meio Ambiente (IEMA) em parceria com o Ministério dos Transportes, cobre o período de 1980 a 2024 e constitui a referência mais abrangente sobre emissões veiculares no Brasil (IEMA; Ministério dos Transportes, 2025). A metodologia é baseada em fatores de emissão por fase PROCONVE e volumes de venda de combustível, gerando estimativas de CO, NO_x, material particulado e CO₂ em nível nacional. A principal diferença em relação ao presente trabalho reside no nível de agregação: o inventário opera com totais nacionais por tipo de veículo e fase regulatória, ao passo que a análise aqui proposta identifica violações individuais por modelo/versão utilizando dados certificados do PBEV, permitindo nomear quais modelos específicos excedem os limites L8.

O CETESB publica anualmente o relatório de qualidade do ar no Estado de São Paulo (CETESB, 2024), que inclui análises de evolução histórica das emissões veiculares. Esses relatórios demonstram a efetividade do PROCONVE a longo prazo, mas não avaliam a conformidade modelo a modelo com a fase mais recente, nem utilizam os dados do PBEV como fonte primária.

Um estudo mais recente, conduzido pela iniciativa TRUE (*Transport Emissions Real-World Monitoring*) em parceria com a Companhia Ambiental do Estado de São Paulo (CETESB) (LEE et al., 2025), mediu em campo, por sensoriamento remoto, as emissões reais de mais de 323 mil veículos na Região Metropolitana de São Paulo entre maio e julho de 2024. Os resultados mostram que veículos leves certificados na fase L3 do PROCONVE (anterior à L8) emitem, em condições reais de uso, até duas vezes mais NO_x, quatro vezes mais CO e dezenove vezes mais hidrocarbonetos do que os respectivos limites de certificação. O estudo evidencia uma lacuna relevante entre emissões de laboratório (certificação) e emissões reais de uso, mas sua abordagem é distinta da deste trabalho: mede emissões em campo por sensoriamento remoto, sem vínculo direto ao modelo/versão certificado, enquanto a presente pesquisa analisa os dados oficiais de certificação do PBEV por modelo/versão para aferir conformidade com os limites da fase mais recente (L8).

A principal lacuna identificada na literatura é a ausência de uma análise de conformidade sistemática e pública, modelo a modelo, com os limites do PROCONVE L8, aplicando corretamente os limites por tipo de combustível e utilizando os dados certificados do PBEV como fonte. Mesmo os estudos mais recentes de sensoriamento remoto (que comparam emissões reais com os limites de certificação) e os de simulação de tráfego (que estimam emissões agregadas por fluxo viário) não realizam essa análise individual e sistemática de conformidade com a fase L8. Este trabalho propõe preencher essa lacuna.

Método

Este capítulo descreve o *pipeline* completo de coleta, extração, tratamento e análise dos dados, bem como as decisões metodológicas que orientaram cada etapa.

3.1 Visão Geral do Método

O Método adotado segue as etapas clássicas de um projeto de análise de dados (TUKEY, 1977), organizadas em cinco estágios sequenciais apresentado na Figura ?? e descritos a seguir.

A primeira etapa consistiu na **identificação e aquisição da fonte de dados**: a tabela PBEV 2026, disponibilizada pelo INMETRO exclusivamente em formato PDF. A escolha dessa fonte foi motivada pelo fato de ser o único repositório oficial e público com dados certificados de emissão de todos os veículos leves comercializados no Brasil, cobrindo 751 modelos/versões de 39 marcas e 16 categorias de veículos.

Na segunda etapa, realizou-se a **extração automatizada**, convertendo as tabelas do PDF em um *dataset* tabular estruturado por meio da biblioteca `pdfplumber` (BIGELOW, 2023). O principal desafio foi o cabeçalho multi-linha com células mescladas, que confunde extratores automáticos; a solução adotada (mapeamento empírico de índices de coluna por verificação com um registro de valor previamente conhecido) é detalhada na Seção 3.3.

A terceira etapa compreendeu o **tratamento e enriquecimento** dos dados, envolvendo limpeza de registros inválidos (linhas de cabeçalho repetidas a cada página do PDF), tipagem de colunas numéricas, normalização de capitalização da coluna `marca` e cálculo de três variáveis derivadas (`co_excesso_pct`, `nmog_excesso_pct` e `viola_l8`) necessárias para a análise de conformidade com os limites do PROCONVE L8, segmentados por tipo de combustível.

Na quarta etapa, produziu-se a **documentação** do *dataset* na forma de um dicionário de variáveis em formato Excel, com três abas: Variáveis (schema completo com tipo e unidade), Legendas do PDF (correspondência entre as notações do documento origi-

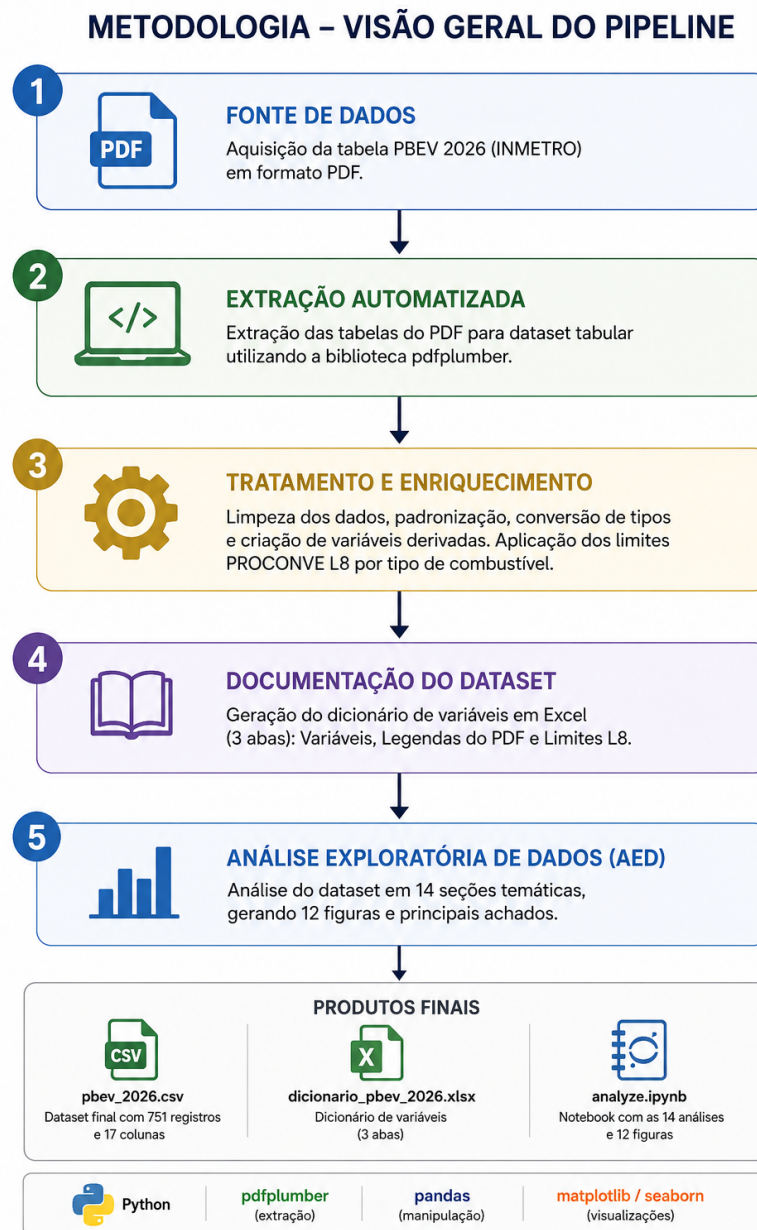


Figura 1 – Fluxo metodológico aplicado nesse trabalho.

nal e os nomes das colunas do CSV) e Limites L8 (parâmetros regulatórios por tipo de combustível).

Por fim, a quinta etapa consistiu na **Análise Exploratória de Dados (AED)**, organizada em 14 seções temáticas com 12 figuras, apresentada no Capítulo 4. Cada seção justifica explicitamente o tipo de visualização adotado, fundamentando as escolhas em Tukey (1977).

Todo o código foi implementado em Python e está organizado em três arquivos principais: `extract_pbev.py` (extração), `make_dict.py` (dicionário) e `analyze.ipynb` (análise exploratória).

3.2 Fonte de Dados

A fonte primária de dados é a **Tabela PBEV 2026, Revisão 04** (janeiro de 2026), publicada pelo INMETRO (INMETRO, 2026). O documento é um PDF de 9 páginas que contém uma tabela com todos os automóveis e comerciais leves certificados para comercialização no Brasil, com informações sobre emissões, consumo energético e classificação ambiental.

3.2.1 Considerações Sobre a Inviabilidade do Uso da FENABRAVE

A proposta inicial deste trabalho incluía cruzar os dados do PBEV com os rankings mensais de vendas da FENABRAVE para ranquear os violadores por popularidade. Essa combinação foi descartada por dois motivos práticos (FENABRAVE, 2025):

A FENABRAVE não disponibiliza rankings de vendas por modelo em formato de dados aberto; os relatórios públicos consolidam os volumes de emplacamentos exclusivamente por fabricante, sem granularidade por modelo. A ANFAVEA, outra fonte relevante de dados de emplacamentos, apresenta a mesma limitação: os dados disponíveis são agrupados por marca, inviabilizando o cruzamento com os registros individuais do PBEV na granularidade necessária a este trabalho.

Essa limitação das fontes de dados de vendas é documentada como limitação deste trabalho (ver Seção 5.3).

3.3 Extração de Dados do PDF

Esta seção detalha o processo metodológico e técnico adotado para a extração e estruturação dos dados brutos contidos no relatório do PBEV 2026. Para garantir a precisão e a reprodutibilidade da coleta, abordam-se a escolha das ferramentas computacionais, o tratamento das complexidades estruturais do documento, como a presença de cabeçalhos multi-linha e células mescladas, e a implementação de filtros dinâmicos de validação de dados.

3.3.1 Biblioteca utilizada

A extração foi realizada com a biblioteca `pdfplumber` (BIGELOW, 2023), que identifica tabelas em PDFs digitalmente gerados a partir da geometria das linhas e da posição dos caracteres. A abordagem é adequada para o PBEV 2026 pois o documento não é escaneado.

A alternativa de usar a API de modelos de linguagem de grande porte (LLMs) para extração foi avaliada e descartada por limitações de *rate limit* de *tokens* de entrada e

saída no nível gratuito, que tornariam o processo impraticável para um documento de 9 páginas com mais de 750 linhas. Essa limitação pode ser mitigada ao longo dos próximos anos.

3.3.2 Problema do cabeçalho multi-linha

A tabela PBEV 2026 possui um cabeçalho multi-linha com células mescladas que confunde os extratores automáticos. Para superar isso, adotou-se uma estratégia de mapeamento empírico de colunas: inspecionou-se a linha correspondente ao modelo **CHEVROLET SPIN 1.8L** (cujo valor de CO = 800 mg/km era previamente conhecido), confirmando cada índice de coluna. Os índices verificados empiricamente são apresentados na Tabela 3.

Tabela 3 – Mapeamento de índices de colunas da tabela PBEV 2026.

Índice	Coluna no <i>dataset</i>	Exemplo (SPIN 1.8L AT)
0	categoria	Médio
1	marca	CHEVROLET
2	modelo	SPIN
3	versao	AT
4	motor	1.8L – 8V
5	tipo_propulsao	Combustão
6	transmissao	A-6
9	combustivel	F
10	nmog_nox_mg_km	46
11	co_mg_km	800 ← confirmado
13	nota_verde_l8	C
15	co2_fossil_g_km	119
23	consumo_mj_km	1,92
25	classificacao_pbe	E

Fonte: Elaborado pelo autor com base na tabela PBEV 2026 (INMETRO, 2026).

3.3.3 Filtro de linhas válidas

Em vez de uma lista de marcas pré-definida (abordagem frágil diante de novas marcas no mercado), o script filtra linhas válidas com três critérios:

Uma linha é aceita somente se satisfizer, simultaneamente, três condições: possuir pelo menos 12 colunas; ter a posição de índice 1 (marca) não-vazia e diferente de palavras-chave de cabeçalho; e ter a posição de índice 2 (modelo) não-vazia. Essa estratégia dispensa uma lista de marcas pré-definida, tornando o extrator robusto a novas entradas no mercado, como as marcas chinesas ZEEKR, LEAPMOTOR e OMODA, presentes no PBEV 2026, que foram capturadas corretamente sem qualquer ajuste manual no código.

3.4 Tratamento e Enriquecimento dos Dados

Após a extração, o *dataset* passou por três etapas de tratamento. O primeiro passo foi a **tipagem**: as colunas numéricas (`co_mg_km`, `nmog_nox_mg_km`, `co2_fossil_g_km` e `consumo_mj_km`) foram convertidas de texto para `float`, com tratamento explícito de valores nulos, que ocorrem naturalmente em veículos elétricos, que não produzem CO ou NMOG+NOx por combustão, e em veículos plug-in, que não registram CO₂ fóssil quando operam em modo elétrico. O segundo passo foi a **normalização** da coluna `marca` para letras maiúsculas, eliminando variações de capitalização entre páginas do PDF. Por fim, calcularam-se as **colunas derivadas** descritas na Seção 3.4.1.

3.4.1 Colunas calculadas

O script `extract_pbev.py` calcula automaticamente três colunas adicionais durante a extração, conforme a Tabela 4.

Tabela 4 – Colunas calculadas no *pipeline* de extração.

Coluna	Fórmula / Descrição
<code>co_excesso_pct</code>	$\max\left(0, \frac{co_mg_km - 700}{700} \times 100\right)$
<code>nmog_excesso_pct</code>	$\max\left(0, \frac{nmog_nox_mg_km - 60}{60} \times 100\right)$
<code>viola_18</code>	True se <code>co_mg_km > 700</code> OU <code>nmog_nox_mg_km > 60</code>

Fonte: Elaborado pelo autor. Nota: os limites de 700 e 60 mg/km são para gasolina/flex. O notebook `analyze.ipynb` recalcula `viola_18` com os limites corretos por combustível.

3.5 Critérios de Conformidade L8 por Combustível

A etapa mais crítica do método é a aplicação dos limites corretos segmentados por tipo de combustível. Conforme a Seção 2.3, os limites do PROCONVE L8 são aplicados de acordo com o tipo de combustível:

$$\text{Gasolina/Flex: } CO \leq 700 \text{ mg/km} \quad \text{e} \quad \text{NMOG+NOx} \leq 60 \text{ mg/km} \quad (1)$$

$$\text{Diesel: } CO \leq 500 \text{ mg/km} \quad \text{e} \quad \text{NMOG+NOx} \leq 460 \text{ mg/km} \quad (2)$$

Os valores substancialmente distintos para diesel refletem as características técnicas dos motores de ciclo Diesel, que operam em regime de mistura pobre e apresentam perfil de emissão diferente dos motores de ciclo Otto.

A análise realizada em `analyze.ipynb` reconstrói o indicador `viola_18` para cada grupo de combustível separadamente. Veículos elétricos (combustível “E”) não têm valores

de CO ou NMOG+NOx e são tratados como automaticamente conformes em relação a esses poluentes.

Na análise preliminar sem segmentação, 63 veículos a diesel foram incorretamente identificados como violadores por ter NMOG+NOx acima de 60 mg/km, valor correto para gasolina/flex, mas muito inferior ao limite diesel de 460 mg/km. Após a segmentação, todos esses veículos são corretamente classificados como conformes.

3.6 Estrutura do Dataset

Antes da apresentação do schema técnico (Tabela 5), é necessário descrever os conceitos associados aos atributos que compõem o *dataset*. Esses atributos podem ser organizados em três grupos conforme sua natureza e função analítica.

O primeiro grupo reúne os **atributos de identificação e caracterização**: categoria de mercado, marca, modelo, versão, motorização, tipo de propulsão, transmissão e combustível. Essas variáveis categóricas permitem segmentar a frota em subgrupos tecnologicamente homogêneos, condição necessária para a aplicação correta dos limites do PROCONVE L8, que variam conforme o tipo de combustível (ver Seção 2.3). O atributo combustível, em particular, é o que determina qual conjunto de limites regulatórios se aplica a cada registro: gasolina e flex seguem os limites mais restritivos de CO (700 mg/km) e NMOG+NOx (60 mg/km), enquanto veículos a diesel seguem limites próprios (500 mg/km de CO e 460 mg/km de NMOG+NOx).

O segundo grupo compreende os **atributos ambientais e energéticos**, que constituem as variáveis numéricas centrais da análise. As emissões de monóxido de carbono (CO, em mg/km) e de hidrocarbonetos não-metânicos somados aos óxidos de nitrogênio (NMOG+NOx, em mg/km) representam os poluentes locais regulados pelo L8 e são as variáveis diretamente comparadas aos limites de conformidade. A emissão de CO₂ fóssil (em g/km) e o consumo energético (em MJ/km) capturam a dimensão climática e de eficiência, respectivamente. Conforme discutido na Seção 2.3, a redução de CO₂ não implica necessariamente a redução de CO ou NMOG+NOx, o que justifica a análise simultânea dessas variáveis para compreender as diferentes dimensões do desempenho ambiental veicular.

O terceiro grupo inclui os **atributos de classificação institucional**: a Nota Verde (A, B ou C), que reflete a avaliação de conformidade ambiental atribuída pelo INMETRO durante a certificação, e a Classificação PBE (A a E), que avalia a eficiência energética. Ambos os indicadores são atribuídos pelo INMETRO com base em critérios de ponderação próprios e permitem confrontar a classificação oficial com os resultados da análise independente conduzida neste trabalho. A coexistência desses dois indicadores na tabela PBEV viabiliza a investigação de se eficiência energética e conformidade de poluentes são métricas correlacionadas ou independentes.

Por fim, o *dataset* contém **atributos derivados**, calculados durante o *pipeline* de tratamento dos dados. O percentual de excesso de CO e de NMOG+NOx em relação aos limites do L8 e o indicador booleano de violação são variáveis que não existem no PDF original e foram construídas para transformar os dados de emissão em indicadores de conformidade diretamente utilizáveis na análise. As fórmulas de cálculo dessas variáveis são detalhadas na Seção 3.4.1.

3.6.1 Pré-processamento da fonte de dados em PDF

A tabela PBEV 2026 é disponibilizada pelo INMETRO em um documento PDF de 9 páginas que, apesar de conter dados precisos e certificados, apresenta uma estrutura que inviabiliza o consumo computacional direto.

O principal obstáculo é o **cabeçalho multi-linha com células mescladas**. Os nomes das colunas estão distribuídos em duas a três linhas do documento, com algumas células abrangendo múltiplas colunas (*column span*) e outras sem rótulo visível, pois o rótulo está na linha imediatamente acima. Extratores automáticos que interpretam o cabeçalho como uma única linha produzem colunas anônimas ou incorretamente atribuídas. Agrava-se o problema pelo fato de que o cabeçalho é repetido no início de cada nova página, introduzindo linhas de cabeçalho espúrias intercaladas aos dados se não forem descartadas.

Um segundo desafio é o **formato numérico brasileiro**: valores decimais com vírgula como separador (por exemplo, “1,92” para 1,92 MJ/km) precisam de conversão explícita para o padrão de ponto flutuante do Python antes de qualquer cálculo. Somam-se ainda os **campos intencionalmente nulos**: veículos elétricos não possuem valores de CO e NMOG+NOx por não realizarem combustão; veículos plug-in (PHEV) não registram CO₂ fóssil quando operam exclusivamente em modo elétrico. Esses nulos representam ausências legítimas de dados, não erros de extração, e exigem tratamento diferenciado.

3.6.2 Dataset resultante e transformações aplicadas

O resultado do processamento é o arquivo `data/pbev_2026.csv`, com 751 linhas e 17 colunas tipadas e documentadas. A transformação central foi realizada pelo script `extract_pbev.py`, que aplicou as seguintes operações sobre o PDF bruto:

A **identificação de colunas por mapeamento empírico** substituiu a extração automática de cabeçalhos. Cada índice de coluna foi verificado linha a linha pelo registro do CHEVROLET SPIN 1.8L (cujo CO = 800 mg/km era previamente conhecido pela Nota Verde C do INMETRO), garantindo correspondência exata entre coluna do CSV e variável do documento original (ver Seção 3.3).

A **conversão de tipos** padronizou valores numéricos para `float` via a função da biblioteca Pandas `pandas.to_numeric(errors='coerce')`, que converte automaticamente

a vírgula decimal e transforma células não-numéricas em NaN sem interromper o processamento. A **normalização de texto** padronizou a coluna **marca** para letras maiúsculas, eliminando inconsistências como “BYD” vs “Byd” em páginas distintas do mesmo documento. O **enriquecimento** adicionou três colunas calculadas (**co_excesso_pct**, **nmog_excesso_pct** e **viola_l8**) que não existem no PDF original e são essenciais para a análise de conformidade.

O dataset resultante é inteiramente reproduzível a partir do PDF original com um único comando: `python extract_pbev.py`. A Tabela 5 apresenta o schema completo, com tipo de dado, unidade e descrição de cada variável. Um dicionário de variáveis detalhado (com três abas: Variáveis, Legendas do PDF e Limites L8) é disponibilizado adicionalmente no arquivo `dicionario_pbev_2026.xlsx`.

Tabela 5 – Schema do *dataset* `pbev_2026.csv`.

Coluna	Tipo	Unidade	Descrição
<code>categoria</code>	texto	—	Segmento de mercado (Sub Compacto, Médio, Picape...)
<code>marca</code>	texto	—	Fabricante (normalizado para maiúsculas)
<code>modelo</code>	texto	—	Nome comercial
<code>versao</code>	texto	—	Trim level (LX, EXL, PREMIER...), 19 nulos
<code>motor</code>	texto	—	Ex.: 1.0-6V, 1.8L-8V, Elétrico
<code>tipo_propulsao</code>	texto	—	Combustão, Elétrico, Híbrido, Plug-in
<code>transmissao</code>	texto	—	M-5, A-6, CVT, DCT-7, N.A....
<code>combustivel</code>	texto	—	F=Flex, G=Gasolina, D=Diesel, E=Elétrico
<code>nmog_nox_mg_km</code>	real	mg/km	NMOG+NOx, 71 nulos (EVs)
<code>co_mg_km</code>	real	mg/km	Monóxido de carbono, 45 nulos (EVs)
<code>nota_verde_l8</code>	texto	—	Conformidade L8 pelo INMETRO: A, B ou C
<code>co2_fossil_g_km</code>	real	g/km	CO ₂ fóssil, 158 nulos (EVs e PHEVs)
<code>consumo_mj_km</code>	real	MJ/km	Consumo energético total, sem nulos
<code>classificacao_pbe</code>	texto	—	Eficiência energética PBE: A (melhor) a E (pior)
<code>co_excesso_pct</code>	real	%	Excesso de CO sobre limite L8 (calculado)
<code>nmog_excesso_pct</code>	real	%	Excesso de NMOG+NOx sobre limite L8 (calculado)
<code>viola_l8</code>	bool	—	True se qualquer limite L8 foi violado

Fonte: Elaborado pelo autor com base na tabela PBEV 2026 (INMETRO, 2026).

3.7 Ferramentas e Reprodutibilidade

A implementação utilizou exclusivamente bibliotecas Python de código aberto. A extração dos dados foi realizada com `pdfplumber` (BIGELOW, 2023), biblioteca que identifica tabelas em PDFs digitalmente gerados a partir da geometria das linhas e da posição

dos caracteres, dispensando o uso de OCR. A manipulação, limpeza e enriquecimento do *dataset* foram feitos com **pandas** (MCKINNEY, 2010), que oferece estruturas de dados tabulares eficientes e expressivas para análise em Python. As visualizações foram produzidas com **matplotlib** e **seaborn**, que juntos cobrem desde gráficos estatísticos simples até painéis compostos com múltiplos subplots.

Todo o processo foi organizado em três arquivos principais: **extract_pbev.py**, responsável pela extração e geração do *dataset*; **make_dict.py**, responsável pela geração do dicionário de variáveis; e **analyze.ipynb**, o notebook Jupyter que concentra as 14 seções da Análise Exploratória de Dados. Essa separação garante que cada etapa possa ser executada, inspecionada e modificada de forma independente.

O código-fonte completo, incluindo os scripts, o *dataset* processado e o notebook de análise, está disponível publicamente em **github.com/felipelemedias/pbev-emissions-pipeline**. O *pipeline* é integralmente reproduzível a partir do PDF original em três comandos:

```
python extract_pbev.py    # gera data/pbev_2026.csv
python make_dict.py       # gera dicionario_pbev_2026.xlsx
jupyter notebook analyze.ipynb  # executa as 14 seções da AED
```

Resultados da Análise Exploratória

Este capítulo apresenta os resultados da Análise Exploratória de Dados (AED) conduzida sobre o *dataset* PBEV 2026. A análise está organizada em 14 seções temáticas, seguindo a estrutura do notebook `analyze.ipynb`, e produz 12 figuras (Figuras 2 a 13).

4.1 Visão Geral do Dataset

O *dataset* final contém 751 registros, isso significa que 100% dos registros disponíveis foram coletados, correspondentes a modelos/versões de veículos leves certificados pelo INMETRO no PBEV 2026. A Tabela 6 apresenta as estatísticas descritivas do conjunto de dados.

Tabela 6 – Estatísticas gerais do *dataset* PBEV 2026.

Métrica	Valor
Total de registros	751
Marcas distintas	39
Categorias de veículos	16
Colunas	17
Período de referência	Janeiro/2026
Nulos em <code>co_mg_km</code>	45 (veículos elétricos)
Nulos em <code>nmog_nox_mg_km</code>	71 (veículos elétricos)
Nulos em <code>co2_fossil_g_km</code>	158 (EVs e PHEVs)

Fonte: Elaborado pelo autor com base na tabela PBEV 2026 (INMETRO, 2026).

4.2 Composição da Frota

Para representar a composição da frota, utilizaram-se **gráficos de pizza** (*pie charts*) para propulsão e combustível, tipo indicado quando o objetivo é mostrar proporções de um todo com até cinco ou seis categorias, pois o ângulo de setor comunica participações relativas de forma imediata. Para a distribuição por categoria de veículo, com 16 grupos,

empregou-se **gráfico de barras horizontais**, pois a comparação visual por comprimento de barra é perceptivamente mais precisa do que por ângulo de setor (TUKEY, 1977). A Figura 2 apresenta os três painéis combinados.

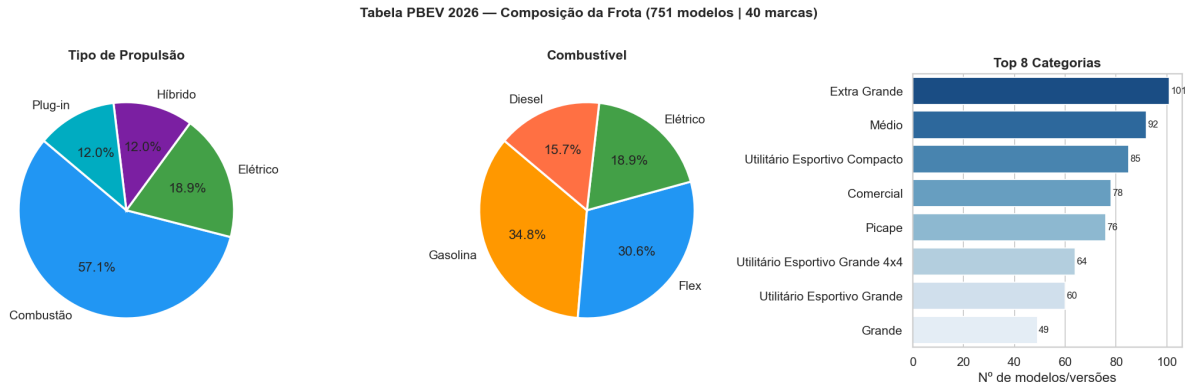


Figura 2 – Composição da frota certificada PBEV 2026 por tipo de propulsão, combustível e categoria de veículo.

A combustão interna ainda responde pela maioria dos modelos (57,1%), mas o avanço dos veículos eletrificados é expressivo: elétricos puros (18,9%), híbridos HEV (11,9%) e plug-in PHEV (11,9%) somam 42,9% dos modelos certificados. A Tabela 7 detalha esses dados.

Tabela 7 – Composição da frota PBEV 2026 por tipo de propulsão e combustível.

Por propulsão			Por combustível	
Tipo	Modelos	%	Combustível	%
Combustão interna	429	57,1%	Gasolina	34,8%
Elétrico puro	142	18,9%	Flex	30,6%
Híbrido (HEV)	89	11,9%	Diesel	15,7%
Plug-in (PHEV)	91	12,1%	Elétrico	18,9%

Fonte: Elaborado pelo autor com base na tabela PBEV 2026 (INMETRO, 2026).

4.3 Conformidade com PROCONVE L8

Para identificar violadores em relação aos limites L8, optou-se pelo **diagrama de dispersão** (*scatter plot*), tipo adequado para visualizar a distribuição conjunta de duas variáveis contínuas e identificar pontos acima de um limiar de referência (TUKEY, 1977). As retas de limite ($\text{CO} = 700 \text{ mg/km}$ e $\text{NMOG} + \text{NOx} = 60 \text{ mg/km}$) são sobrepostas como referências visuais diretas; a alternativa de dois histogramas independentes não permitiria detectar violações simultâneas de ambos os poluentes. A Figura 3 apresenta o diagrama segmentado por grupo de combustível.

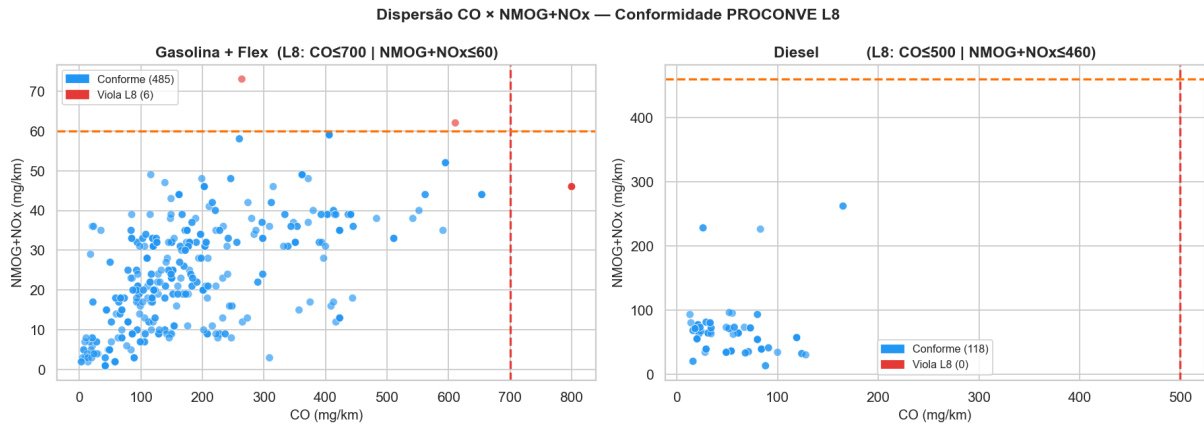


Figura 3 – Dispersão CO $\times$ NMOG+NO $_x$ com limites PROCONVE L8 por combustível. Pontos à direita ou acima das linhas de limite indicam violação.

A análise foi conduzida em dois grupos:

Gasolina + Flex (limites: CO $\leq$ 700 mg/km e NMOG+NO $_x \leq$ 60 mg/km): dos 491 modelos nesse grupo, a grande maioria está em conformidade. Ao todo, **seis modelos/versões violam os limites** (1,2% do grupo): cinco versões do CHEVROLET SPIN 1.8L e o GWM TANK 300 PHEV.

Diesel (limites: CO $\leq$ 500 mg/km e NMOG+NO $_x \leq$ 460 mg/km): após a aplicação dos limites corretos para diesel, nenhum veículo desse grupo viola os limites do L8. A análise bruta (sem segmentação) havia identificado erroneamente 63 veículos diesel como violadores por aplicar os limites de gasolina/flex.

Esse resultado demonstra a importância metodológica da segmentação por tipo de combustível, conforme discutido na Seção 3.5.

4.4 Distribuição de CO: Gasolina e Flex

Para revelar a forma da distribuição de CO, empregou-se um **histograma**, representação que divide o domínio em intervalos de largura uniforme e exibe a frequência de observações em cada intervalo, evidenciando assimetria, moda e a posição do limite L8 em relação à massa de dados (TUKEY, 1977). O histograma foi preferido ao diagrama de caixa (*boxplot*) como visualização principal porque o objetivo desta seção é a forma completa da distribuição, não apenas o sumário de cinco estatísticas. Um **diagrama de caixa** complementar é utilizado ao lado para a comparação compacta entre os grupos flex e gasolina. A Figura 4 apresenta ambos.

A distribuição de CO é assimétrica à direita em ambos os grupos: a maior parte dos modelos modernos apresenta emissões abaixo de 400 mg/km, com poucos valores acima do limite. A Tabela 8 resume as estatísticas.

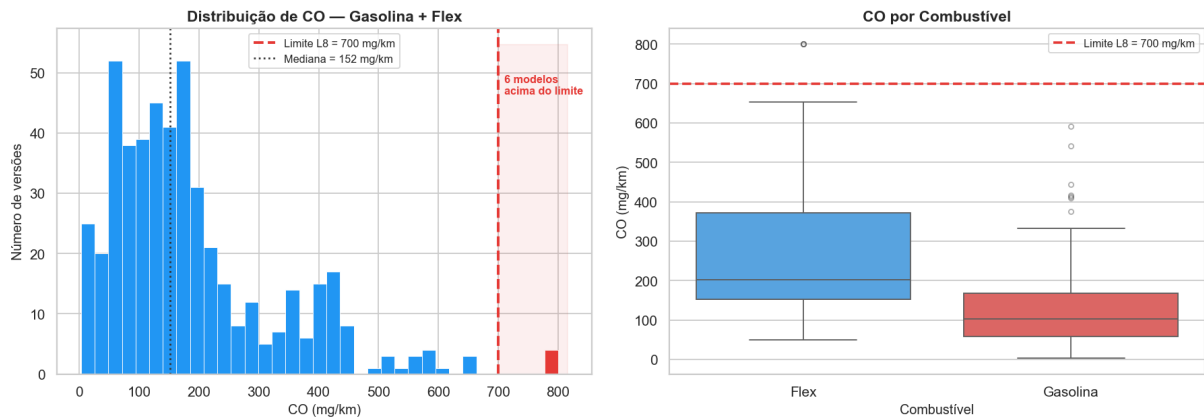


Figura 4 – Distribuição de CO (mg/km) para veículos a gasolina e flex. A linha vermelha indica o limite L8 de 700 mg/km.

Tabela 8 – Estatísticas de CO (mg/km) para gasolina e flex.

Combustível	Mín.	Mediana	Média	Máx.
Gasolina	0	≈105	≈120	≈265
Flex	0	≈200	≈230	800

Fonte: Elaborado pelo autor com base na tabela PBEV 2026 (INMETRO, 2026).

A mediana de CO para veículos flex é significativamente maior do que para gasolina. Isso se deve à presença de modelos com motorizações mais antigas (como o SPIN 1.8L) na base flex, que foram desenvolvidos antes da vigência do L8. Os veículos a gasolina puros tendem a ser modelos importados ou de marcas que já adotaram plataformas mais modernas.

4.5 Distribuição de NMOG+NOx

O **histograma** foi mantido para NMOG+NOx pelos mesmos motivos que para CO: a distribuição é fortemente concentrada em valores baixos, com a maioria dos modelos abaixo de 30 mg/km, e o histograma revela essa concentração e a posição isolada dos violadores de forma mais direta que um boxplot (TUKEY, 1977). A Figura 5 apresenta a distribuição com o limite de 60 mg/km destacado.

A média de NMOG+NOx é de aproximadamente 30 mg/km, bem abaixo do limite de 60 mg/km, demonstrando que a maior parte da frota a gasolina/flex já estava em conformidade com esse parâmetro antes mesmo da entrada em vigor do L8. Os poucos modelos que superam o limite são motorizações antigas de motor 6V (ex.: SPIN 1.8L, versão manual).

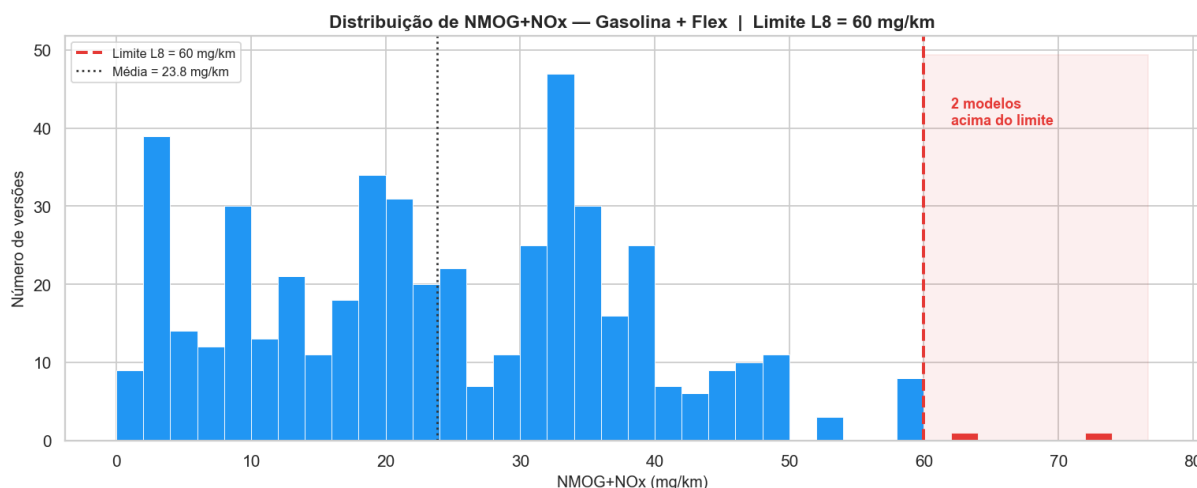


Figura 5 – Distribuição de NMOG+NOx (mg/km) para gasolina e flex, com limite L8 de 60 mg/km indicado.

4.6 Top 20 Maiores Emissores de CO

O **gráfico de barras horizontais ordenadas** é a representação canônica para rankings: a orientação horizontal acomoda rótulos de modelos longos sem necessidade de rotação, e a ordenação decrescente torna o pior caso imediatamente visível (TUKEY, 1977). A Figura 6 apresenta os 20 modelos com maior emissão de CO (mg/km) na base PBEV 2026.

O **CHEVROLET SPIN 1.8L** domina as primeiras posições com emissão de CO de 800 mg/km (excesso de +14,3% sobre o limite de 700 mg/km). Quatro versões do SPIN (AT, LT automático, LTZ e PREMIER) apresentam esse valor. O quinto maior emissor de CO é o **RENAULT DUSTER 1.6L**, com 654 mg/km (dentro do limite). A partir da 5ª posição, todos os modelos estão dentro do limite de 700 mg/km.

A Tabela 9 apresenta o ranking final dos violadores de CO e/ou NMOG+NOx entre veículos a gasolina e flex.

4.7 Análise por Marca

Para a análise por marca, utilizaram-se dois **gráficos de barras horizontais** independentes, um para taxa de não-conformidade e outro para CO médio, pois as escalas são incompatíveis (porcentagem *vs.* mg/km) e um gráfico de eixo duplo distorceria a percepção visual das magnitudes (TUKEY, 1977). A Figura 7 apresenta o resultado segmentado por gasolina e flex.

Na análise bruta (sem segmentação por combustível), marcas com grande proporção de veículos diesel apresentavam taxas de violação artificialmente elevadas. Após a correção, **duas marcas** apresentam modelos violadores dos limites de gasolina/flex: a

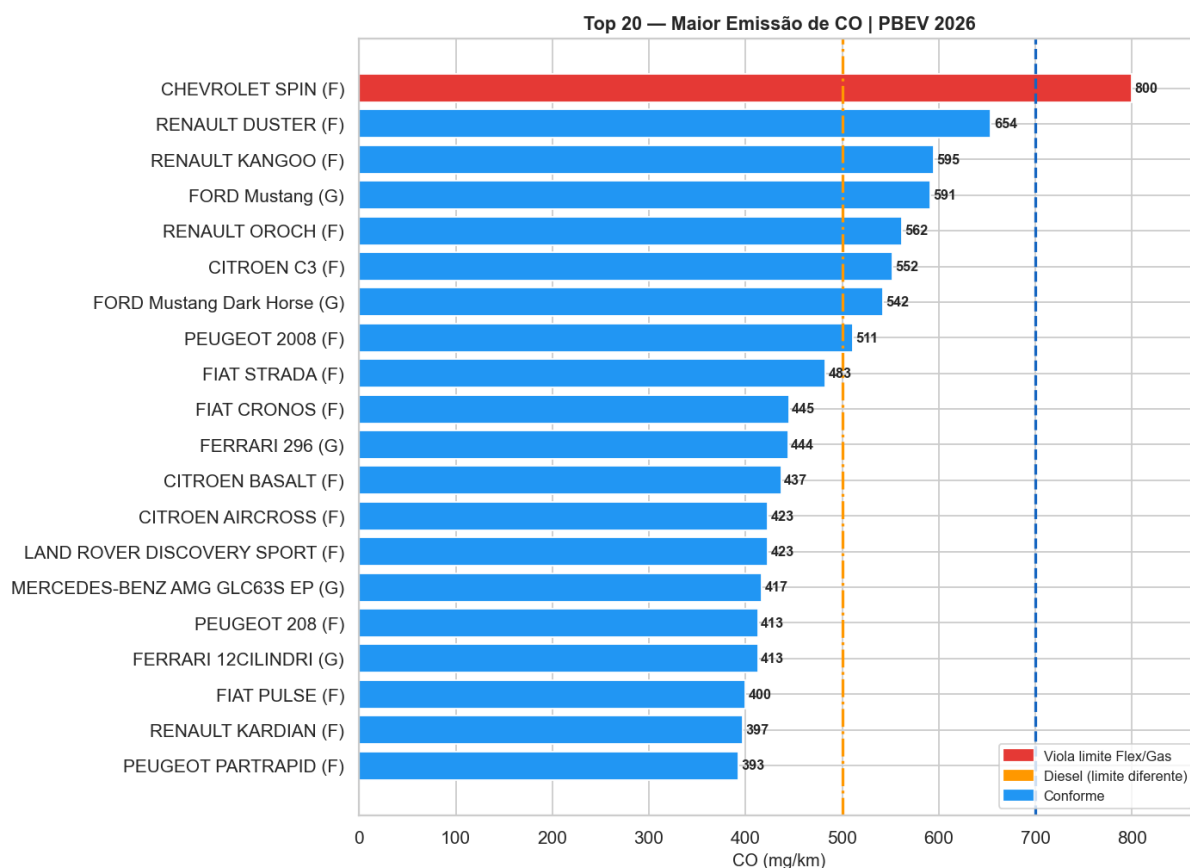


Figura 6 – Top 20 modelos com maior emissão de CO (mg/km). A linha vermelha indica o limite L8 de 700 mg/km para gasolina/flex.

CHEVROLET (cinco versões do SPIN 1.8L) e a **GWM** (TANK 300 PHEV, violação de NMOG+NOx). As demais marcas têm todas as versões em conformidade plena com o L8.

4.8 Diagnóstico de Categoria

A mesma estratégia de dois **gráficos de barras horizontais** independentes foi adotada aqui pelos mesmos motivos de incompatibilidade de escala (TUKEY, 1977). A ordenação das barras por taxa de não-conformidade facilita a leitura direta de quais categorias concentram mais modelos fora do L8. A Figura 8 apresenta o número de violadores e o CO mediano por categoria.

4.8.1 Distribuição da frota por categoria

As 16 categorias do PBEV 2026 agrupam os veículos por porte e função. As categorias de maior volume de modelos certificados são Sub Compacto, Compacto e SUV, concentrando a maior parte dos modelos a gasolina e flex destinados ao mercado de massa.

Tabela 9 – Ranking de violadores do PROCONVE L8 entre veículos a gasolina e flex (6 versões/modelos).

#	Marca	Modelo	Versão	CO (mg/km)	Exc. CO	NMOG+NO _x	Nota Verde
1	CHEVROLET	SPIN	AT	800	+14,3%	46 mg/km	C
2	CHEVROLET	SPIN	LT (aut.)	800	+14,3%	46 mg/km	C
3	CHEVROLET	SPIN	LTZ	800	+14,3%	46 mg/km	C
4	CHEVROLET	SPIN	PREMIER	800	+14,3%	46 mg/km	C
5	CHEVROLET	SPIN	LT (man.)	611	— (ok)	62 mg/km	C
6	GWM	TANK 300	PHEV	264	— (ok)	73 mg/km	B

Fonte: Elaborado pelo autor com base na tabela PBEV 2026 (INMETRO, 2026). Notas: (1) o SPIN LT manual viola NMOG+NO_x ($62 > 60$, excesso de +3,3%) mas não CO ($611 < 700$); (2) o GWM TANK 300 PHEV viola NMOG+NO_x ($73 > 60$, excesso de +21,7%) mas não CO ($264 < 700$); a Nota Verde B (e não C) reflete a ponderação interna do INMETRO sobre o conjunto de todos os poluentes regulados.

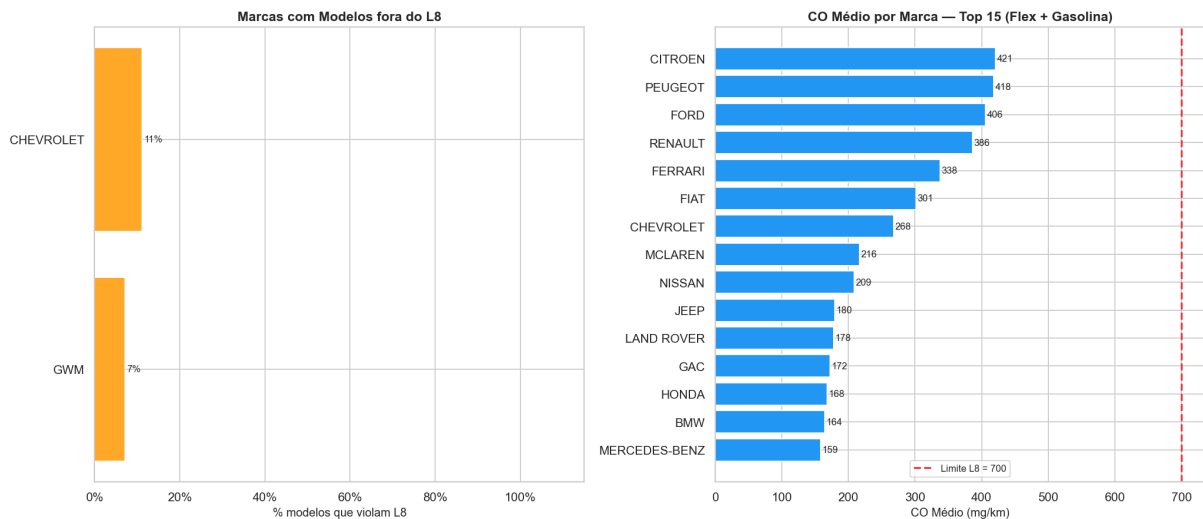


Figura 7 – Taxa de violação e emissão média de CO por marca, segmentado por combustível.

Categorias como Picape, Comercial e Utilitário Leve reúnem proporcionalmente mais veículos a diesel, o que explica a distorção observada na análise bruta de conformidade (ver Seção 4): o alto número de falsos positivos nessas categorias decorre exclusivamente da aplicação incorreta dos limites de gasolina/flex aos motores diesel.

4.8.2 Categorias e perfil de emissão

Após a segmentação por combustível, o perfil de emissão de CO por categoria torna-se mais nítido. As categorias que concentram modelos com motorizações maiores e mais antigas, como **Médio** e **Minivan**, apresentam medianas de CO mais elevadas no grupo gasolina/flex. É nessas categorias que se encontram as cinco versões do SPIN 1.8L não-

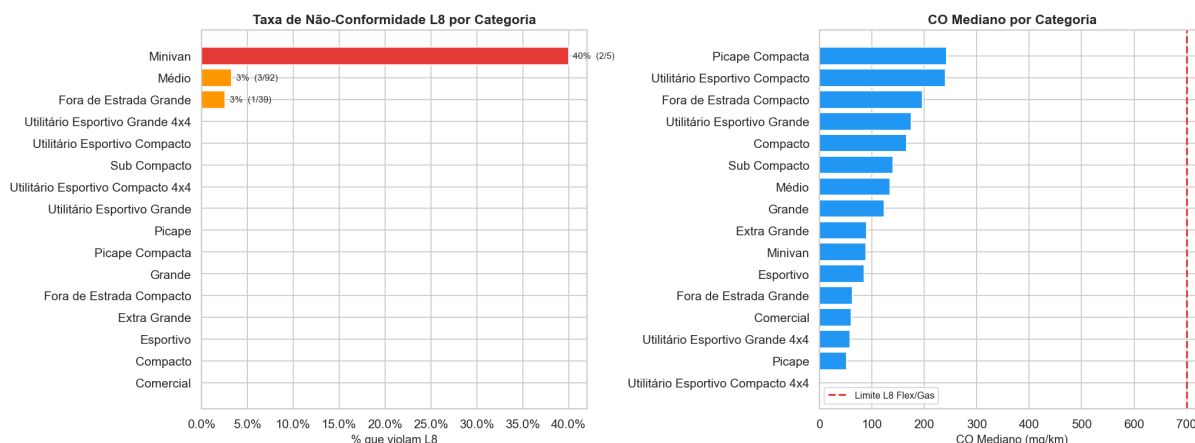


Figura 8 – Taxa de Não Conformidade PROCONVE L8 e CO mediano por categoria de veículo.

conformes: quatro na subcategoria Médio (versões AT, LT automático, LTZ e PREMIER) e uma na configuração de Minivan. Categorias dominadas por veículos importados modernos (como SUV de alto padrão e EVs) apresentam as menores medianas de CO, refletindo plataformas desenvolvidas já sob regimes regulatórios mais restritivos.

Os violadores residuais de gasolina/flex concentram-se, portanto, em categorias de veículos de médio porte com histórico de motorização nacional, onde a atualização de plataformas de motor é mais lenta em função do volume de produção e da cadeia de fornecimento estabelecida.

4.8.3 Implicações regulatórias por categoria

A distribuição dos violadores por categoria é relevante para a formulação de políticas de renovação de frota. Programas de incentivo à substituição de veículos com motorizações legadas tendem a ter maior impacto sobre as categorias onde os ciclos de atualização de plataforma são mais longos. No contexto brasileiro, a categoria Médio combina alta penetração em frotas corporativas e de aplicativo com um ciclo de renovação tecnológica mais lento do que o de veículos sub-compactos, em parte porque o volume de produção nacional cria inércia contra a descontinuação de linhas rentáveis.

A presença de um veículo plug-in (GWM TANK 300 PHEV) entre os violadores na categoria SUV aponta para um desafio emergente: veículos com tração híbrida plug-in possuem motor a combustão que opera em modo gerador quando a bateria está descarregada, produzindo emissões em condições de carga distintas das de um veículo de combustão convencional. O ciclo de teste do PBEV pode capturar o motor em condições operacionais às quais o sistema de pós-tratamento não foi plenamente calibrado.

4.9 Diagnóstico Ambiental

Os resultados de conformidade apresentados nas seções anteriores possuem implicações que transcendem a simples contagem de modelos violadores. Esta seção examina o significado sanitário, regulatório e socioeconômico das emissões identificadas.

4.9.1 Significância dos poluentes identificados para a saúde

O monóxido de carbono (CO) é reconhecido pela Organização Mundial da Saúde (OMS) como um poluente prioritário em ambientes urbanos (WHO, 2021). Em concentrações elevadas, o CO compromete a capacidade de transporte de oxigênio no sangue ao formar carboxiemoglobina, com efeitos que vão de sintomas leves (cefaleia, tontura) a danos neurológicos e cardiovasculares em exposições prolongadas. Motores em combustão incompleta, condição frequente em arranques a frio, acelerações bruscas e congestionamentos, são a principal fonte veicular de CO em áreas urbanas.

Os hidrocarbonetos não-metânicos (NMOG) identificados na violação da versão manual do SPIN 1.8L são precursores de ozônio troposférico e de material particulado secundário. Compostos específicos do grupo NMOG, como benzeno e 1,3-butadieno, são classificados como carcinogênicos pela Agência Internacional de Pesquisa em Câncer (IARC) (WHO, 2021). Em regiões metropolitanas com alta densidade veicular e baixa ventilação, essas emissões contribuem significativamente para episódios de *smog* fotoquímico e para o aumento de internações por doenças respiratórias (CETESB, 2024).

4.9.2 Contexto da conformidade no primeiro ano do L8

O PBEV 2026 representa o primeiro ciclo de certificação inteiramente sob os critérios da Fase L8, vigente desde janeiro de 2025. Nesse contexto, a taxa de conformidade de 98,2% no segmento gasolina/flex indica que a indústria automotiva, em geral, concluiu as adaptações tecnológicas antes da entrada em vigor do novo limite. A violação concentrada em uma única plataforma de motor (bloco 1.8L 8V do SPIN) é indicativa de uma exceção tecnológica pontual, e não de uma falha sistêmica de adequação do setor.

Do ponto de vista regulatório, veículos com Nota Verde C estão expostos a restrições crescentes à medida que políticas de renovação de frota e zonas de baixas emissões (ZBE) avançam no Brasil. São Paulo e outras capitais já discutem instrumentos que limitam a circulação de veículos com pior desempenho ambiental em períodos de má qualidade do ar (CETESB, 2024).

4.9.3 Impactos além do domínio ambiental

As emissões veiculares têm repercussões em outros domínios além do meio ambiente.

Saúde pública: a exposição crônica ao CO e ao NMOG em ambientes urbanos está associada ao aumento de internações por doenças respiratórias e cardiovasculares, com impacto direto nos sistemas de saúde. A magnitude desse efeito depende do volume de veículos violadores em circulação, dado não disponível neste trabalho por ausência de dados de emplacamentos por modelo (ver Seção 5.3).

Mercado e valor de revenda: veículos com Nota Verde C tendem a encontrar maior resistência de mercado à medida que consumidores e gestores de frotas corporativas incorporam critérios ambientais em suas decisões de compra. A Nota Verde, visível na etiqueta PBE exposta no veículo, funciona como sinalizador direto ao consumidor final.

Responsabilidade do fabricante: a manutenção de modelos com emissões acima dos limites legais no catálogo ativo expõe os fabricantes a riscos regulatórios e reputacionais. Para a CHEVROLET, o elemento técnico central da não-conformidade do SPIN 1.8L é a plataforma de motor 8V legada; a atualização para injeção eletrônica de maior precisão e catalisador de três vias de maior eficiência seria o caminho natural. Para a GWM, o GWM TANK 300 PHEV apresenta $\text{NMOG} + \text{NOx} = 73 \text{ mg/km}$ em modo gasolina (excesso de +21,7%), indicando que o sistema de pós-tratamento de hidrocarbonetos não-metânicos pode não estar totalmente calibrado para o ciclo de teste brasileiro.

4.9.4 Mobilidade urbana, qualidade do ar e zonas de baixas emissões

As emissões identificadas neste trabalho têm implicações diretas sobre políticas de mobilidade urbana em curso no Brasil. O conceito de **zonas de baixas emissões (ZBE)**, áreas urbanas onde a circulação de veículos com piores desempenhos ambientais é restrin-gida ou tributada, já está sendo debatido em São Paulo e outras capitais brasileiras com referência explícita à Nota Verde do INMETRO (CETESB, 2024). Veículos com Nota Verde C, como as cinco versões do SPIN 1.8L identificadas, estariam sujeitos a restrições nesse tipo de zona, com impactos concretos sobre sua utilidade para proprietários em áreas metropolitanas.

A qualidade do ar em ambientes urbanos de alta densidade, medida por índices como o IQAr (Índice de Qualidade do Ar) monitorado pelo CETESB (CETESB, 2024), é diretamente afetada pelas emissões de CO e $\text{NMOG} + \text{NOx}$. Embora seis modelos/versões violadores representem 1,2% do segmento gasolina/flex do PBEV 2026, o impacto real sobre o IQAr depende do volume absoluto desses veículos em circulação nas regiões metropolitanas, dado não disponível neste trabalho por ausência de dados de emplacamentos por modelo (ver Seção 5.3).

4.9.5 Custos ao sistema de saúde pública

A exposição crônica a CO e NMOG+NO_x em ambientes urbanos está associada ao aumento de internações hospitalares por doenças respiratórias e cardiovasculares. Estudos de valoração econômica da poluição do ar estimam custos ao sistema público de saúde na ordem de bilhões de reais por ano apenas no Estado de São Paulo (CETESB, 2024). Embora este trabalho não realize tal valoração (a ausência de dados de frota circulante por modelo torna qualquer extrapolação de custo especulativa), o resultado de conformidade apresentado fornece o insumo necessário: a identificação de quais veículos emitem acima do limite e em quanto.

Apesar deste trabalho não apresentar nenhuma análise histórica sobre a quantidade de veículos em circulação e sobre o mercado de cada fabricante, a ideia é, no primeiro momento, de demonstrar os dados de forma individualizada.

O raciocínio de cadeia de causalidade é o seguinte: modelos violadores em circulação → excedente de CO e NMOG em regiões metropolitanas → aumento marginal da concentração atmosférica desses poluentes → incremento nas internações por doenças respiratórias e cardiovasculares → custo ao SUS. Cada elo dessa cadeia requer dados adicionais, mas a identificação dos modelos violadores, o primeiro elo, é exatamente o que este trabalho contribui.

4.10 CO₂ Fóssil e Consumo Energético por Tecnologia

Para comparar CO₂ fóssil e consumo por tecnologia, empregou-se o **diagrama de caixa** (*boxplot*) para o CO₂, tipo que mostra mediana, quartis e valores extremos e é robusto a distribuições assimétricas. Para o consumo energético, utilizou-se o **gráfico de violino** (*violin plot*), que combina o boxplot com uma estimativa da densidade da distribuição, útil aqui para revelar a bimodalidade nos veículos híbridos (HEV), que não seria capturada por um boxplot simples (TUKEY, 1977). A Figura 9 apresenta ambas as visualizações.

Os resultados, sumarizados na Tabela 10, revelam que os veículos plug-in reduzem o CO₂ fóssil em $\approx 70\%$ em relação à combustão convencional (34 vs. 111 g/km na mediana). Os veículos elétricos eliminam totalmente o CO₂ fóssil direto. Híbridos HEV apresentam ganho modesto em CO₂ (103 g/km) mas eficiência energética sensivelmente melhor (1,84 MJ/km vs. 1,95 MJ/km da combustão).

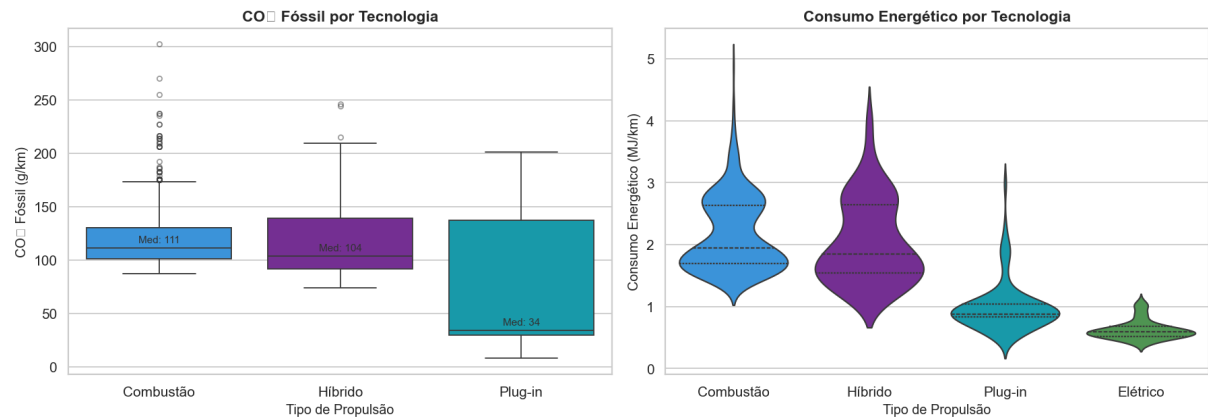


Figura 9 – CO₂ fóssil (g/km) e consumo energético (MJ/km) medianos por tipo de propulsão.

Tabela 10 – CO₂ fóssil e consumo energético medianos por tipo de propulsão.

Tecnologia	CO ₂ fóssil mediano (g/km)	Consumo mediano (MJ/km)
Combustão interna	111	1,95
Híbrido (HEV)	103	1,84
Plug-in (PHEV)	34	0,88
Elétrico puro	0	0,59

Fonte: Elaborado pelo autor com base na tabela PBEV 2026 (INMETRO, 2026).

4.11 Nota Verde INMETRO × Conformidade Calculada

Para a tabulação cruzada, empregou-se um **mapa de calor anotado** (*annotated heatmap*), tipo recomendado para matrizes de contingência porque a codificação por cor identifica rapidamente as células dominantes e as anotações numéricas garantem precisão sem recorrer à legenda de cor (TUKEY, 1977). O **gráfico de barras agrupadas** ao lado complementa com a distribuição proporcional de Nota Verde por tipo de propulsão. A Figura 10 apresenta os dois painéis.

A Tabela 11 apresenta a tabulação cruzada. O resultado mais relevante é que os 63 veículos com Nota Verde A marcados como violadores na análise bruta são *todos diesel*, pois o INMETRO os classificou corretamente com Nota A porque aplicou os limites diesel. Após a segmentação, emergem dois padrões:

A **Nota C** concentra os violadores mais expressivos: os 5 SPINs são os únicos com essa nota que violam efetivamente os limites de gasolina/flex. A **Nota B** revela um achado inesperado: o GWM TANK 300 PHEV tem Nota Verde B atribuída pelo INMETRO e, ainda assim, viola NMOG+NO_x (73 mg/km, +21,7%). Isso indica que o INMETRO, ao calcular a Nota Verde, aplica uma ponderação que considera todos os poluentes regula-

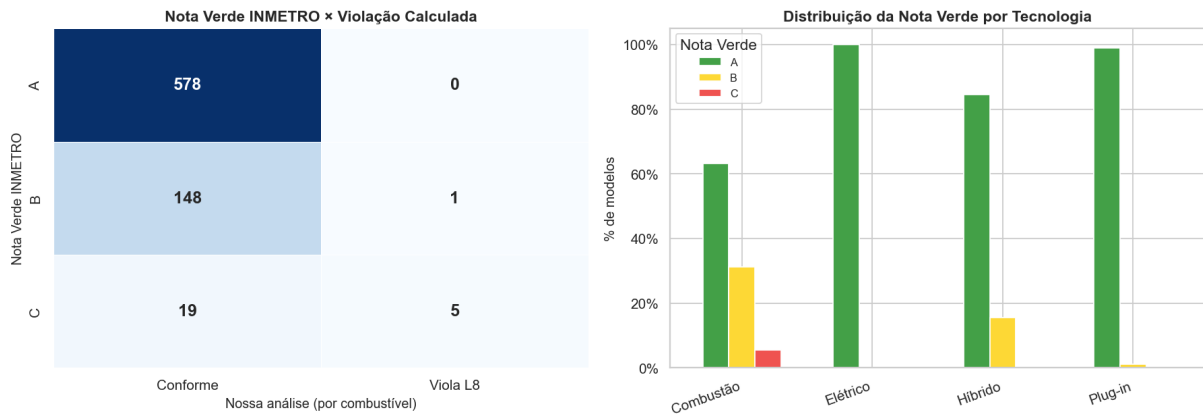


Figura 10 – Distribuição de Nota Verde INMETRO versus conformidade calculada com limites L8 segmentados por combustível.

dos (CO, NMOG, NO_x isolado, MP, THC), e o GWM pode ser suficientemente bom nos demais para não atingir o nível C, apesar do excesso de NMOG+NO_x.

Tabela 11 – Tabulação cruzada: Nota Verde INMETRO versus conformidade calculada com limites L8 (análise segmentada por combustível).

Nota Verde	Conformes (nossa análise)	Violam (nossa análise)
A	578 (após segmentação diesel)	0
B	148	1 (GWM TANK 300 PHEV†)
C	19	5 (SPIN 1.8L, gasolina/flex)

Fonte: Elaborado pelo autor. †O GWM TANK 300 PHEV viola NMOG+NO_x (73 > 60 mg/km) mas recebeu Nota B (não C) do INMETRO, refletindo ponderação sobre o conjunto completo de poluentes regulados.

4.12 Classificação PBE × Conformidade L8

Para analisar a relação entre classificação PBE (variável ordinal de cinco níveis: A a E) e emissões, empregou-se o **diagrama de caixa por grupo**, tipo que permite identificar tendências monotônicas entre classes ordenadas ao mesmo tempo que mostra a dispersão interna de cada grupo (TUKEY, 1977). A Figura 11 apresenta dois painéis: CO por classe PBE e consumo energético por classe PBE.

O resultado confirma que **eficiência energética (PBE) e conformidade de poluentes (L8) são métricas independentes**. Um veículo pode ter PBE = E (menos eficiente energeticamente) e ainda assim ser plenamente conforme com os limites de poluentes do L8. O SPIN 1.8L é um caso em que ambos os critérios convergem negativamente: PBE = E e violação do L8, mas por razões técnicas distintas. Modelos elétricos são os

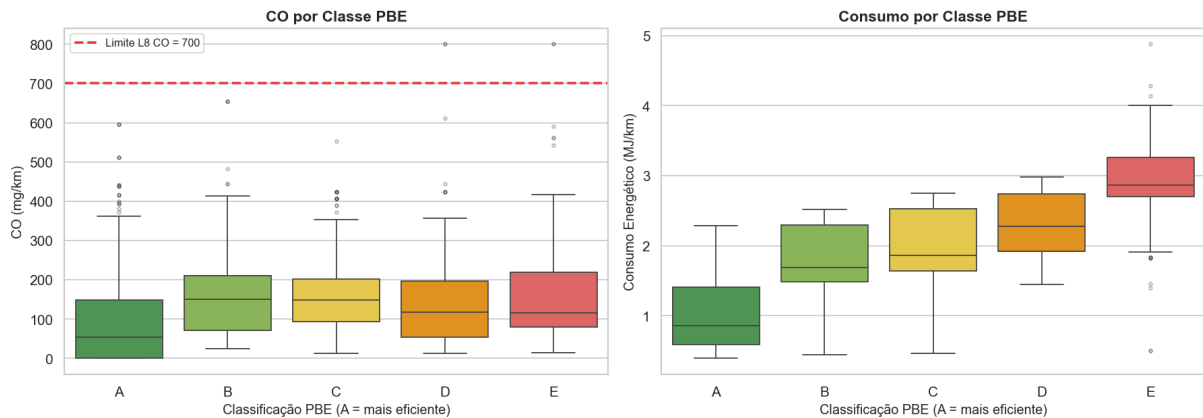


Figura 11 – Classificação PBE versus emissão de CO e consumo energético por classe PBE.

únicos que maximizam simultaneamente PBE = A, Nota Verde = A e emissões locais nulas.

4.13 Correlações entre Variáveis

4.13.1 Matriz de correlação geral: todos os pares

Para representar as correlações, empregou-se um **mapa de calor de correlação** (*correlation heatmap*): visualização compacta que codifica simultaneamente a direção (azul = negativa, vermelho = positiva) e a magnitude de todos os pares por meio de gradiente de cor (TUKEY, 1977). A Figura 12 apresenta a matriz completa.

A Tabela 12 apresenta os seis pares de variáveis e seus coeficientes de Pearson (r), calculados sobre as linhas com valores válidos em ambas as variáveis.

A estrutura da matriz revela dois agrupamentos distintos. O **cluster de eficiência energética** (CO_2 fóssil consumo, $r = 0,90$) é fortemente coeso. O **cluster de poluentes locais** (CO, NMOG+NOx) correlaciona-se moderadamente com o cluster de eficiência (r entre 0,37 e 0,42), mas CO e NMOG+NOx praticamente não se correlacionam entre si ($r = 0,11$). Essa estrutura sugere que eficiência de combustão e qualidade de pós-tratamento são dimensões relativamente independentes do desempenho ambiental de um veículo.

4.13.2 Correlações de senso comum

Uma análise crítica requer confrontar as expectativas intuitivas, baseadas no senso comum técnico e prático, com os resultados observados nos dados. No contexto da engenharia automotiva e da física, o senso comum refere-se ao conjunto de premissas intuitivas e simplificadas derivadas do conhecimento geral sobre o funcionamento dos motores a com-

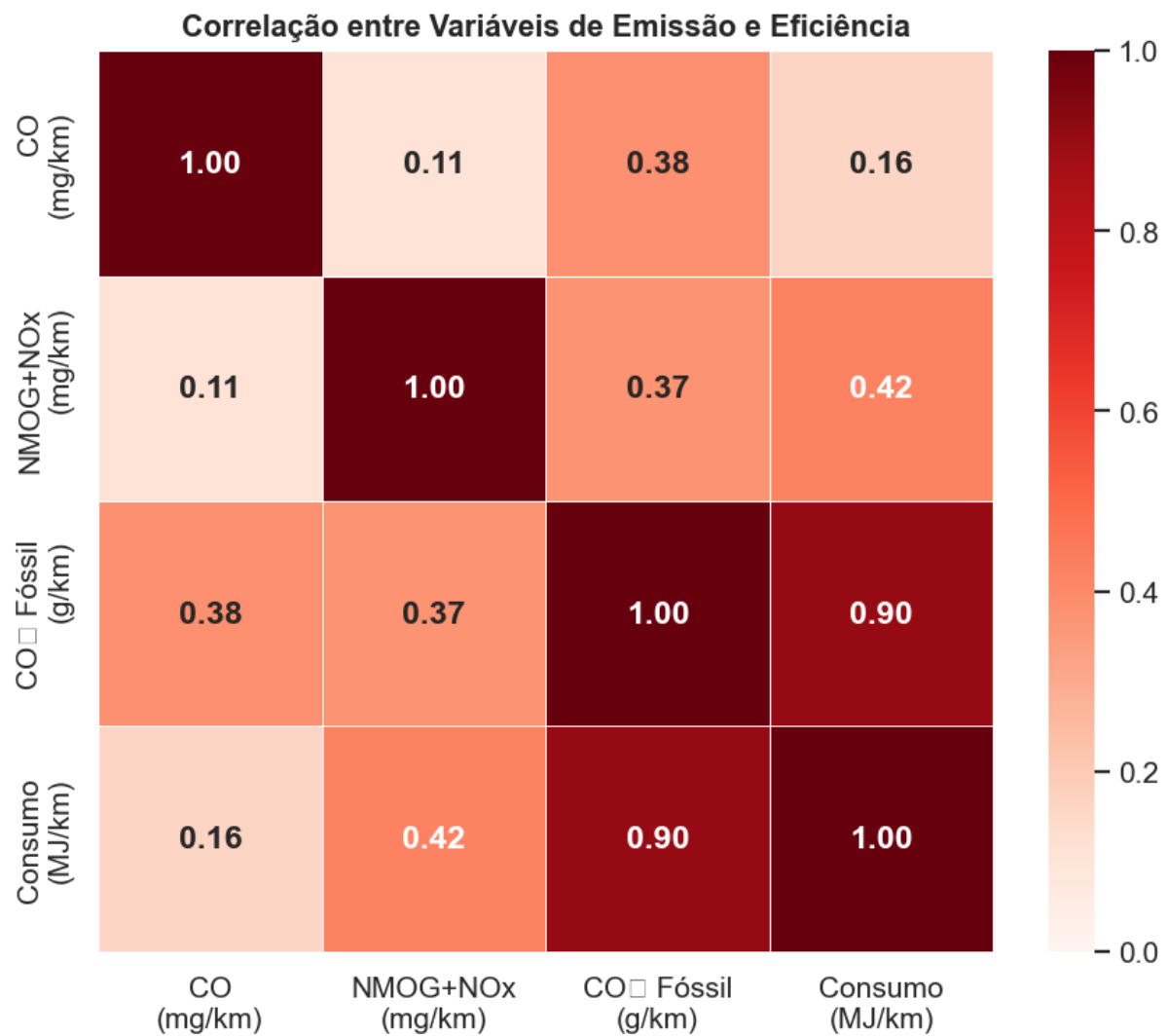


Figura 12 – Mapa de calor de correlação de Pearson entre as variáveis numéricas do PBEV 2026.

bustão interna (por exemplo, a noção empírica de que maior queima de combustível gera proporcionalmente maior volume de todos os tipos de emissões) (FERGUSON; KIRKPATRICK, 2015). Longe de ser um mero pré-conceito, esse saber intuitivo serve como uma importante referência inicial (ou hipótese nula de trabalho) na análise de dados: é a partir do estabelecimento desse *baseline* intuitivo que se torna possível validar a coerência do *dataset* com a teoria fundamental ou, de forma ainda mais relevante, identificar desvios, anomalias e desacoplamentos tecnológicos complexos que desafiam as expectativas simplificadas.

CO₂ × consumo (r = 0,90, confirmada): esta é a correlação de maior “senso comum” do dataset e a que apresenta o mecanismo causal mais direto. CO₂ é um produto direto e quase linear da combustão de combustíveis fósseis: mais combustível queimado implica mais CO₂ emitido, por lei da termodinâmica. A força $r = 0,90$ confirma essa

Tabela 12 – Correlações de Pearson entre todos os pares de variáveis numéricas do PBEV 2026.

Par de variáveis	r	Interpretação
CO ₂ fóssil × Consumo MJ/km	0,90	Forte positiva: combustão produz CO proporcionalmente ao combustível
CO ₂ fóssil × Consumo MJ/km	0,90	Forte positiva: combustão produz CO proporcionalmente ao combustível
NMOG+NO _x × Consumo MJ/km	0,42	Moderada positiva: maior consumo associa-se a mais emissões de HC
CO × CO ₂ fóssil	0,38	Moderada positiva: combustão como causa comum
NMOG+NO _x × CO ₂ fóssil	0,37	Moderada positiva: mesma causa comum da combustão
CO × Consumo MJ/km	0,16	Fraca: CO independe da quantidade de combustível
CO × NMOG+NO _x	0,11	Quase nula: controlados independentemente por catalisadores modernos

Fonte: Elaborado pelo autor com base na tabela PBEV 2026 (INMETRO, 2026). Correlações calculadas descartando pares com valor nulo.

expectativa.

NMOG+NO_x × consumo (r = 0,42; moderada, plausível): maior consumo energético indica motores maiores ou mais carregados, que tendem a produzir mais hidrocarbonetos não-metânicos em valor absoluto mesmo com catalisadores eficientes. A correlação moderada é tecnicamente coerente.

NMOG+NO_x × CO₂ (r = 0,37; moderada, causa comum): hidrocarbonetos não-metânicos e CO₂ correlacionam-se moderadamente não porque um cause o outro, mas porque ambos derivam da combustão de combustíveis fósseis, a variável latente comum.

CO × CO₂ e CO × consumo (r = 0,38 e 0,16; fracas a moderadas): a intuição “mais combustível queimado → mais CO” é apenas parcialmente confirmada. CO é produto da *qualidade* da combustão (incompletude da reação), não da *quantidade* de combustível. Um motor de alta cilindrada com catalisador de três vias eficiente emite pouco CO apesar do alto consumo.

CO × NMOG+NO_x (r = 0,11; quase nula, contra-intuitiva): este é o achado mais relevante do ponto de vista analítico. Apesar de ambos serem produtos de combustão incompleta, CO e NMOG+NO_x têm correlação praticamente zero. A explicação está nos sistemas de pós-tratamento modernos: o catalisador de três vias controla CO e NMOG de formas parcialmente independentes, respondendo de maneira diferente à temperatura, ao envelhecimento e à composição do combustível. No dataset PBEV 2026, há modelos com CO alto e NMOG baixo (SPIN 1.8L AT: CO = 800, NMOG+NO_x = 46) e modelos com CO normal e NMOG alto (GWM TANK 300: CO = 264, NMOG+NO_x = 73). A quase-independência desses dois poluentes implica que instrumentos regulatórios precisam monitorá-los separadamente, o que o PROCONVE L8 já faz ao estabelecer limites

distintos para cada parâmetro.

4.13.3 Análise de causalidade

Aplicando o raciocínio de correlação *vs.* causalidade ao conjunto completo dos seis pares de variáveis do PBEV 2026 e seguindo o trabalho de Kašpar, Fornasiero e Hickey (2003):

CO₂ ↔ consumo (r = 0,90), causalidade direta: queimar combustível fóssil produz CO₂ por estequiometria da reação de oxidação: mais combustível implica mais CO₂, sem intermediários nem variáveis de confundimento relevantes.

NMOG+NOx ↔ consumo (r = 0,42), causalidade mediada: motores maiores ou mais carregados produzem mais hidrocarbonetos em valor absoluto, mesmo com catalisadores eficientes. A correlação é real, mas mediada pela eficiência do pós-tratamento.

CO ↔ CO₂ e NMOG+NOx ↔ CO₂, causa comum: CO e NMOG+NOx correlacionam-se moderadamente com CO₂ não porque um cause o outro, mas porque todos derivam da combustão. É um caso clássico de correlação por variável latente.

CO ↔ NMOG+NOx (r = 0,11), quase nula, desacoplamento técnico: os sistemas de pós-tratamento modernos controlam CO e NMOG de formas tecnicamente desacopladas, produzindo os dois casos tecnicamente ortogonais de violação identificados neste trabalho.

A implicação prática é direta: **políticas de redução de CO₂ (metas climáticas) não garantem redução automática de CO e NMOG+NOx (metas de saúde pública)**. São dimensões distintas que requerem instrumentos regulatórios específicos, justificando a coexistência do PROCONVE (poluentes locais) e das metas de eficiência energética do PBE (CO₂ e consumo).

4.14 Síntese dos Achados

O **painel de síntese** (*summary card*) corresponde à etapa de síntese narrativa que encerra toda AED, conforme proposto por Tukey (1977). Diferentemente dos gráficos estatísticos convencionais, o painel consolida os principais achados em linguagem direta, com codificação por cor por categoria temática, visando tornar os resultados acessíveis a públicos não especializados. A Figura 13 apresenta esse painel.

Os principais achados da AED são:

1. 751 modelos/versões de 39 marcas foram certificados no PBEV 2026, cobrindo toda a frota de veículos leves novos no Brasil.
2. A frota é majoritariamente a combustão (57%), mas veículos eletrificados já somam 43% dos modelos certificados.

Síntese dos Principais Achados — EDA PBEV 2026

DATASET	751 modelos/versões · 40 marcas · Tabela PBEV 2026 (INMETRO Jan/2026)
FROTA	57% combustão interna · 19% elétrico puro · 12% híbrido · 12% plug-in
L8 FLEX/GAS	6 de 491 modelos Flex/Gasolina violam PROCONVE L8 (1,2%)
PIOR CASO	CHEVROLET SPIN 1.8L — CO = 800 mg/km (+14,3% acima do limite de 700 mg/km)
DIESEL	Limites L8 diesel diferentes (CO≤500, NMOG+NOx≤460) — 0 diesel violam
MARCAS	RAM e CITROEN com maiores taxas de violação; RENAULT com CO persistentemente alto
ELÉTRICOS	142 EVs puros: CO _{fóssil} = 0 · Nota Verde A · Consumo ~0.5 MJ/km
CO _{fóssil}	Plug-in reduz CO _{fóssil} ~69% vs combustão (111 → 34 g/km)
PBE × L8	Eficiência energética (PBE) é INDEPENDENTE de conformidade L8 — critérios distintos

Figura 13 – Painel de síntese com os principais achados da análise exploratória do PBEV 2026.

Fonte: Elaborado pelo autor com base na tabela PBEV 2026 (INMETRO, 2026).

- Os limites L8 são combustível-específicos. Analisar diesel com limites de gasolina/flex gera 63 falsos positivos.
- Entre veículos flex e gasolina, **6 modelos/versões violam o L8** (1,2%): 5 versões do CHEVROLET SPIN 1.8L e o GWM TANK 300 PHEV.
- Quatro versões do SPIN violam CO com 800 mg/km (+14,3%); a versão manual viola NMOG+NOx com 62 mg/km (+3,3%).
- O GWM TANK 300 PHEV viola NMOG+NOx com 73 mg/km (+21,7%); sua Nota Verde B reflete a ponderação INMETRO sobre o conjunto completo de poluentes.
- A Nota Verde INMETRO é consistente com a análise após a segmentação por combustível.
- Eficiência energética (PBE) e conformidade de poluentes (L8) são métricas independentes.
- CO e NMOG+NOx têm correlação quase nula ($r = 0,11$): sistemas de pós-tratamento os controlam de forma tecnicamente desacoplada.

10. Veículos plug-in reduzem CO₂ fóssil em $\approx 70\%$; elétricos eliminam totalmente o CO₂ fóssil direto.

4.15 Análise de Cenários

Esta seção explora cenários hipotéticos derivados dos resultados encontrados, com o objetivo de dimensionar a sensibilidade das conclusões e estimar impactos adicionais não diretamente mensuráveis pela contagem de modelos violadores. Os cenários são construídos sobre as quatro variáveis numéricas do dataset (CO, NMOG+NO_x, CO₂ fóssil e consumo energético), abordando tanto os limites regulatórios (CO e NMOG+NO_x) quanto as implicações climáticas e energéticas (CO₂ e consumo).

4.15.1 Cenário 1: Aplicação de um limite de CO mais restritivo

O limite de CO do PROCONVE L8 para gasolina/flex (700 mg/km) já representa uma redução de 46% em relação ao L7 (1.300 mg/km). Se o limite fosse reduzido para 650 mg/km, aproximando-se de padrões internacionais mais restritivos, o 5º maior emissor da base PBEV 2026, o RENAULT DUSTER 1.6L (654 mg/km), passaria à condição de não-conforme, elevando o número de modelos violadores de CO no grupo gasolina/flex de 4 para pelo menos 5. A margem entre esse modelo e o limite é de apenas 4 mg/km, equivalente a 0,6% acima do hipotético novo limite.

Esse cenário ilustra que a franja próxima ao limite vigente é estreita: poucos modelos se encontram na zona de risco imediato. A adoção de limites progressivamente mais restritivos em fases futuras (L9, L10) deve ampliar essa pressão sobre motorizações com tecnologia de pós-tratamento menos eficiente.

4.15.2 Cenário 2: Estimativa de excedente de CO por quilometragem

As quatro versões do SPIN 1.8L que violam o limite de CO emitem, por km percorrido, 100 mg além do permitido. Considerando uma quilometragem anual de 20.000 km, referência típica para veículos de uso misto urbano/rodoviário no Brasil (CETESB, 2024), o excedente de CO por veículo por ano é:

$$\Delta\text{CO}_{\text{anual}} = 100 \text{ mg/km} \times 20.000 \text{ km/ano} = 2,0 \text{ kg de CO/ano por veículo} \quad (3)$$

Para a versão manual, que viola o NMOG+NO_x (62 mg/km, excesso de 2 mg/km):

$$\Delta\text{NMOG+NO}_x_{\text{anual}} = 2 \text{ mg/km} \times 20.000 \text{ km/ano} = 40 \text{ g de NMOG+NO}_x\text{/ano por veículo} \quad (4)$$

A ausência de dados de emplacamentos por modelo (limitação discutida na Seção 5.3) impede a extrapolação desses valores para a frota total em circulação. No entanto, as equações acima fornecem um indicador por unidade que pode ser aplicado por analistas com acesso aos dados de volume de vendas.

4.15.3 Cenário 3: Conformidade total com atualização dos modelos violadores

Os seis modelos/versões não-conformes representam 1,2% do grupo gasolina/flex (6/491). Se as cinco versões do SPIN 1.8L tivessem suas plataformas de motor atualizadas (ou descontinuas) e o GWM TANK 300 PHEV tivesse o sistema de pós-tratamento de NMOG recalibrado para o ciclo brasileiro, a taxa de conformidade do segmento gasolina/flex passaria de 98,8% para 100%.

Esse cenário confirma que a não-conformidade identificada é **pontual e tecnologicamente localizada em dois modelos de perfis distintos**: uma motorização a combustão legada (SPIN 1.8L) e um veículo plug-in de alto padrão (TANK 300 PHEV).

4.15.4 Cenário 4: Endurecimento do limite de NMOG+NOx

O limite atual de NMOG+NOx para gasolina/flex no L8 é de 60 mg/km. Se esse limite fosse reduzido para 55 mg/km, redução de 8,3% compatível com a trajetória de aperto entre o L7 (80 mg/km) e o L8, novos modelos entrariam na condição de não-conformidade.

A versão manual do SPIN 1.8L (62 mg/km) continuaria violando o novo limite com folga. O GWM TANK 300 PHEV (73 mg/km) igualmente. O cenário mais significativo seria o surgimento de novos violadores borderline: modelos na faixa atual de 55–59 mg/km que são conformes ao L8 vigente mas passariam a violar o hipotético L9.

4.15.5 Cenário 5: Impacto de deslocamento da frota para tecnologias de baixa emissão

Este cenário estima o impacto potencial sobre as variáveis de CO₂ fóssil e consumo energético se uma fração da frota atual a combustão migrasse para tecnologias eletrificadas, com base nas medianas observadas no dataset PBEV 2026 (ver Tabela 10).

Se 10% dos 429 modelos de combustão interna (43 modelos) fossem substituídos por equivalentes plug-in (PHEV), a redução média de CO₂ por modelo seria da ordem de:

$$\Delta\text{CO}_{2\text{plug-in}} = (111 - 34) \text{ g/km} = 77 \text{ g/km por veículo} \quad (5)$$

Aplicada a uma quilometragem anual de 20.000 km, a redução por veículo seria de:

$$\Delta\text{CO}_{2\text{anual por veículo}} = 77 \text{ g/km} \times 20.000 \text{ km} = 1,54 \text{ kg de CO}_2/\text{ano} \quad (6)$$

Para uma frota de 43 veículos (10% da combustão certificada): 66,2 kg de CO₂/ano, numericamente modesto quando expresso em unidades de certificação (modelos), mas potencialmente significativo quando extrapolado para a frota em circulação, que inclui múltiplas unidades de cada modelo.

Conclusão

Este trabalho investigou a conformidade dos veículos leves comercializados no Brasil em 2026 com os limites de emissão do PROCONVE Fase L8, utilizando como fonte de dados a tabela PBEV 2026 (INMETRO). O percurso metodológico envolveu o desenvolvimento de um *pipeline* de extração automatizada de dados de PDF, a construção de um *dataset* estruturado, a documentação das variáveis e a condução de uma Análise Exploratória de Dados (AED) em 14 seções temáticas.

5.1 Cumprimento dos Objetivos

Objetivo geral: o trabalho identificou os modelos violadores, quantificou o excesso percentual e categorizou os resultados por marca, combustível e categoria. A pergunta de pesquisa foi respondida de forma direta e documentada.

Objetivos específicos:

- ❑ A extração automatizada via `pdfplumber` foi bem-sucedida, convertendo 9 páginas de PDF em 751 registros estruturados com precisão verificada empiricamente;
- ❑ O *dataset* `pbev_2026.csv` e o dicionário `dicionario_pbev_2026.xlsx` foram construídos e documentados com 17 variáveis;
- ❑ Os limites L8 foram aplicados com segmentação por combustível, eliminando os 63 falsos positivos gerados pela análise bruta;
- ❑ Os percentuais de excesso foram calculados e incorporados ao *dataset*;
- ❑ A AED produziu 12 visualizações temáticas abrangendo composição da frota, conformidade, distribuições de poluentes, análise por marca/categoria, CO₂, Nota Verde, PBE e correlações;

- A conformidade calculada foi comparada com a Nota Verde INMETRO e a classificação PBE, confirmando a consistência da certificação oficial após a segmentação correta.

5.2 Principais Contribuições

Contribuição metodológica: a identificação e tratamento do problema da segmentação por combustível na análise de conformidade com o PROCONVE L8 é a principal contribuição metodológica deste trabalho. A análise demonstra que aplicar os limites de gasolina/flex a veículos a diesel gera mais de 60 falsos positivos, um erro sistemático que poderia comprometer qualquer análise de conformidade baseada em dados do PBEV.

Contribuição técnica: o *pipeline* de extração automatizada do PDF do PBEV e o *dataset* resultante são disponibilizados como subprodutos da pesquisa. Esses artefactos permitem que outros pesquisadores conduzam análises sobre a frota certificada sem necessidade de reprocessar o documento PDF original.

Contribuição analítica: a análise demonstra que a frota de veículos leves certificada em 2026 está amplamente em conformidade com o PROCONVE L8. Apenas dois fabricantes apresentam modelos violadores no segmento gasolina/flex: a CHEVROLET, com as cinco versões do SPIN 1.8L (motorização legada, bloco 8V anterior ao L8), e a GWM, com o TANK 300 PHEV ($\text{NMOG} + \text{NOx} = 73 \text{ mg/km}$ em modo gasolina, excesso de +21,7%). A análise causal dos seis pares de variáveis numéricas distingue três tipos de relação: causalidade direta ($\text{CO}_2 \leftrightarrow \text{consumo}$, $r = 0,90$), causalidade mediada ($\text{NMOG} + \text{NOx} \leftrightarrow \text{consumo}$, $r = 0,42$) e correlação por causa comum (CO e $\text{NMOG} + \text{NOx}$ correlacionados com CO_2 por terem a combustão como variável latente). O achado mais relevante é a correlação quase nula entre CO e $\text{NMOG} + \text{NOx}$ ($r = 0,11$), evidenciando desacoplamento técnico no pós-tratamento, com implicações diretas para o desenho de políticas regulatórias.

5.3 Limitações

As principais limitações deste trabalho são descritas a seguir.

Dados de vendas por modelo indisponíveis em formato aberto. A FENABRAVE e a ANFAVEA disponibilizam volumes de emplacamentos apenas por fabricante, sem granularidade por modelo. Isso impossibilita ranquear os violadores por popularidade ou estimar o impacto real dos modelos não-conformes na qualidade do ar, pois o número de unidades em circulação é desconhecido.

Análise restrita ao ano 2026. O trabalho é de corte transversal e cobre apenas a edição PBEV 2026 (Revisão 04, janeiro de 2026). Sem o cruzamento com edições anteriores (2025, 2024), não é possível avaliar a evolução temporal da conformidade nem detectar a trajetória de adequação dos modelos violadores.

Coluna de NOx isolado não extraída. O PDF do PBEV 2026 apresenta ambiguidade na posição da coluna de NOx (isolado de NMOG). A análise ficou restrita a CO e NMOG+NOx; o poluente NOx isolado, cujo limite L8 é de 0,06 g/km (gasolina/flex), não foi avaliado individualmente.

Subcobertura do `pdfplumber` em quebras de página. A extração automática perdeu aproximadamente 18 linhas de dados por incompatibilidade entre quebras de página do PDF e as configurações padrão da biblioteca. Isso representa subcobertura de $\approx 2\%$ no *dataset*. Para trabalhos futuros, recomenda-se ajustar os parâmetros de `table_settings` da biblioteca para mitigar essa perda.

Ciclo de teste nacional (não WLTP). Os valores de emissão certificados no PBEV são medidos pelo ciclo brasileiro, que difere do ciclo WLTP adotado na Europa. Comparações diretas com normas internacionais requerem cautela, pois os perfis de velocidade e carga são distintos.

5.4 Trabalhos Futuros

A partir dos resultados e limitações identificadas, propõem-se as seguintes direções para pesquisas futuras:

- ❑ **Análise temporal:** comparar a conformidade L8 entre as edições PBEV 2025 e 2026 para verificar a evolução da frota após a entrada em vigor da nova fase;
- ❑ **Integração com dados de vendas:** caso fontes de dados de emplacamentos por modelo se tornem disponíveis em formato aberto, seria possível estimar o impacto real dos violadores na qualidade do ar em termos de volume de veículos comercializados;
- ❑ **Extração de NOx:** aprimorar o *pipeline* de extração para capturar a coluna de NOx (isolado de NMOG) da tabela PBEV, permitindo uma análise mais granular da conformidade;
- ❑ **Análise preditiva:** desenvolver modelos para prever a conformidade de novos modelos com base em características técnicas (motorização, propulsão, categoria);
- ❑ **Comparação internacional:** comparar os resultados da frota brasileira certificada sob o L8 com frotas certificadas sob normas equivalentes em outros países (EURO 6, EPA Tier 3).

5.5 Considerações Finais

O PROCONVE L8 representa um avanço regulatório significativo na redução de poluentes locais no Brasil. Os dados do PBEV 2026 mostram que a indústria automobilística,

em sua grande maioria, já opera em conformidade com os novos limites. As exceções identificadas, o CHEVROLET SPIN 1.8L e o GWM TANK 300 PHEV, representam perfis tecnológicos distintos de não-conformidade: o primeiro reflete a permanência de uma motorização legada; o segundo indica que veículos plug-in de alto padrão também podem apresentar emissões de NMOG+NOx acima dos limites em modo gasolina, o que reforça a necessidade de monitoramento contínuo e modelo a modelo.

A disponibilidade pública dos dados do PBEV, aliada a ferramentas de extração automatizada, oferece um valioso recurso para monitoramento contínuo da conformidade ambiental da frota nacional. Espera-se que os artefatos produzidos neste trabalho (*dataset*, dicionário e código de extração) contribuam para ampliar o uso analítico dessas informações por pesquisadores, consumidores e formuladores de políticas públicas.

Referências

BIGELOW, J. **pdfplumber: Plumb a PDF for detailed information about each char, rectangle, and line**. 2023. GitHub. Acesso em: 27 abr. 2026. Disponível em: <<https://github.com/jsvine/pdfplumber>>.

CARVALHO, C. H. R. d. **Emissões relativas de poluentes do transporte motorizado de passageiros nos grandes centros urbanos brasileiros**. Brasília, 2011. Acesso em: 20 jun. 2026. Disponível em: <<https://repositorio.ipea.gov.br/items/4461522e-dd50-4347-beb6-ed5d012f89f4>>.

CETESB. **Relatório Anual de Qualidade do Ar no Estado de São Paulo 2023**. São Paulo, 2024. Acesso em: 27 abr. 2026. Disponível em: <<https://cetesb.sp.gov.br/ar/publicacoes-relatorios/>>.

CONAMA. **Resolução nº 18, de 6 de maio de 1986 – Dispõe sobre a criação do Programa de Controle da Poluição do Ar por Veículos Automotores – PROCONVE**. Brasília, 1986. Acesso em: 27 abr. 2026.

FENABRAVE. **Relatório de Emplacamentos de Veículos – Dados Mensais**. [S.l.], 2025. Acesso em: 31 mar. 2025. Disponível em: <<https://www.fenabrave.org.br>>.

FERGUSON, C. R.; KIRKPATRICK, A. T. **Internal combustion engines: applied thermosciences**. [S.l.]: John Wiley & Sons, 2015.

IBAMA. **PROCONVE – Programa de Controle da Poluição do Ar por Veículos Automotores**. [S.l.], 2024. Acesso em: 27 abr. 2026. Disponível em: <<https://www.gov.br/ibama/pt-br/assuntos/emissoes/proconve>>.

IEMA; Ministério dos Transportes. **Inventário Nacional de Emissões Atmosféricas por Veículos Automotores Rodoviários 2025**. Brasília, 2025. Acesso em: 20 jun. 2026. Disponível em: <<https://www.gov.br/transportes/pt-br/assuntos/sustentabilidade/mudanca-do-clima/inventario-nacional-de-emissoes-atmosfericas-por-veiculos-automotores-rodoviarios-2025>>.

INMETRO. **PBE Veicular – Programa Brasileiro de Etiquetagem Veicular**. [S.l.], 2024. Acesso em: 27 abr. 2026. Disponível em: <<https://www.gov.br/inmetro/pt-br/assuntos/veiculos/pbe-veicular>>.

_____. **Tabela PBEV 2026 – Programa Brasileiro de Etiquetagem Veicular: Automóveis e Comerciais Leves, Revisão 04**. Brasília, 2026. Acesso em: 20 jan. 2026.

JACONDINO, G. B. **Quantificação das Emissões Veiculares Através do Uso de Simuladores de Tráfego**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 2005. Acesso em: 20 ago. 2026. Disponível em: <<http://www.producao.ufrgs.br/arquivos/publicacoes/GabrielBittencourtJacondino.pdf>>.

KAŠPAR, J.; FORNASIERO, P.; HICKEY, N. Automotive catalytic converters: current status and some perspectives. **Catalysis today**, Elsevier, v. 77, n. 4, p. 419–449, 2003.

LEE, K. et al. **Assessment of Real-World Vehicle Emissions in São Paulo**. [S.l.], 2025. Acesso em: 20 ago. 2026. Disponível em: <<https://trueinitiative.org/research/assessment-of-real-world-vehicle-emissions-in-sao-paulo/>>.

LOPES, R. d. S. et al. Análise da eficiência energética da frota de veículos leves no brasil. In: **Anais do XXV Simpósio Internacional de Engenharia Automotiva (SIMEA)**. São Paulo: Blucher, 2017. Acesso em: 20 jun. 2026. Disponível em: <<https://www.proceedings.blucher.com.br/article-details/anlise-da-eficincia-energtica-da-frota-de-veculos-leves-no-brasil-26572>>.

MCKINNEY, W. Data structures for statistical computing in Python. In: WALT, S. van der; MILLMAN, J. (Ed.). **Proceedings of the 9th Python in Science Conference**. [S.l.: s.n.], 2010. p. 51–56.

RAPP, V. et al. Lean-burn internal combustion engines. **Lean combustion**, Elsevier, p. 111–146, 2016.

TUKEY, J. W. **Exploratory Data Analysis**. Reading, MA: Addison-Wesley, 1977.

WHO. **WHO Global Air Quality Guidelines: Particulate Matter, Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide**. Geneva, 2021. Acesso em: 27 abr. 2026. Disponível em: <<https://www.who.int/publications/i/item/9789240034228>>.