

# Introdução a Ciência de Dados

**Projeto:**  
Qualidade do Ar na América Latina

---

## Componentes:

Carolina Penido Barcellos  
Gabrielly Xavier dos Santos  
Marcelo Henrique Pereira Oliveira e Silva  
Matheus Soares dos Santos de Freitas

---

Dezembro 2025

## Sumário

<b>1</b>	<b>Introdução</b>	<b>3</b>
1.1	Estrutura do Relatório . . . . .	4
<b>2</b>	<b>Fundamentação Teórica</b>	<b>4</b>
2.1	Poluentes Atmosféricos Analisados . . . . .	4
2.1.1	Material Particulado Fino (PM <sub>2.5</sub> ) . . . . .	4
2.1.2	Material Particulado Grosso (PM <sub>10</sub> ) . . . . .	4
2.1.3	Dióxido de Nitrogênio (NO <sub>2</sub> ) . . . . .	5
2.1.4	Ozônio Troposférico (O <sub>3</sub> ) . . . . .	5
2.1.5	Dióxido de Enxofre (SO <sub>2</sub> ) . . . . .	5
2.1.6	Monóxido de Carbono (CO) . . . . .	5
2.2	Diretrizes da OMS (2021) . . . . .	5
2.3	Mecanismos de Dispersão Atmosférica . . . . .	6
<b>3</b>	<b>Dados e Metodologia</b>	<b>6</b>
3.1	Base de Dados . . . . .	6
3.1.1	Padronização e Pré-Processamento . . . . .	7
3.2	Análise Exploratória de Dados (EDA) . . . . .	7
3.3	Testes de Hipóteses . . . . .	7
3.3.1	Teste F de Fisher para Igualdade de Variâncias . . . . .	7
3.3.2	Teste de Welch . . . . .	8
3.4	Correlação e Modelagem . . . . .	8
3.4.1	Modelo Preditivo para PM <sub>2.5</sub> . . . . .	8
3.4.2	Modelo Explicativo para PM <sub>2.5</sub> . . . . .	9
3.5	Ferramentas Computacionais . . . . .	10
3.6	Limitações da Análise . . . . .	10
<b>4</b>	<b>Resultados</b>	<b>10</b>
4.1	Estatísticas Descritivas e Distribuições . . . . .	10
4.1.1	Avaliação de Capitais Via Métricas de Qualidade do Ar . . . . .	13
4.2	Padrões Sazonais e Mensais . . . . .	16
4.2.1	Calendários de Calor (Heatmaps) . . . . .	18
4.3	Comparação Entre Capitais . . . . .	19
4.4	Resultados dos Testes de Hipótese . . . . .	20
4.5	Correlação Entre Variáveis . . . . .	22
4.5.1	Correlação Entre Gases e Variáveis Meteorológicas . . . . .	22
4.5.2	Correlação Entre os Gases . . . . .	23

4.6	Modelos Preditivos . . . . .	24
4.6.1	Preparação das Features . . . . .	25
4.6.2	Estrutura do Modelo . . . . .	25
4.6.3	Resultados . . . . .	25
4.6.4	Interpretação . . . . .	26
4.6.5	Importância das Variáveis por Cidade . . . . .	26
4.7	Modelos Explicativos e Inferência Estatística . . . . .	27
4.7.1	Estratégia Geral de Modelagem . . . . .	27
4.7.2	Abordagem 1: LASSO com Validação Cruzada (LassoCV) . . .	28
4.7.3	Abordagem 2: Grid Search com Multiprocessing . . . . .	28
4.7.4	Abordagem 3: Busca Bayesiana com Optuna . . . . .	29
<b>5</b>	<b>Discussão</b>	<b>30</b>
<b>6</b>	<b>Conclusão</b>	<b>30</b>
<b>7</b>	<b>Referências</b>	<b>31</b>
<b>Apêndice A – Testes de Igualdade de Variâncias</b>		<b>31</b>
<b>Apêndice B – Gráficos Complementares</b>		<b>32</b>

## 1 Introdução

Este relatório tem como objetivo apresentar e discutir os resultados obtidos a partir da análise de dados climáticos e de qualidade do ar das capitais latino-americanas, no contexto da realização da COP30 (Conferência das Partes), sediada em Belém, Pará, Brasil. A proximidade desse evento internacional reforça a importância de compreender a posição do Brasil e de seus vizinhos continentais frente aos desafios climáticos contemporâneos, identificando fragilidades regionais, avanços já observados e potenciais pontos de atenção para políticas públicas. Para isso, buscamos responder às seguintes perguntas:

- Como a qualidade do ar em Brasília se compara com a das capitais de outros países como Chile, Colômbia ou Argentina?
- Quais variáveis meteorológicas (temperatura, velocidade do vento, pressão atmosférica) são os preditores mais importantes para a qualidade do ar, e essa importância varia entre as diferentes regiões ?
- É possível desenvolver um modelo de regressão que preveja os níveis de PM2.5 para cidades da América Latina, usando dados meteorológicos e históricos como entrada?
- A média dos níveis de poluição em Brasília difere de forma estatisticamente significativa da média das outras capitais da América Latina?

A base de dados utilizada passou por processos de tratamento e padronização e foi posteriormente analisada sob três eixos principais:

1. **Análise exploratória:** identificação de padrões temporais, sazonalidades e distribuições das concentrações de poluentes;
2. **Testes de hipótese:** investigação de diferenças estatisticamente significativas entre níveis médios de poluição das capitais latino-americanas e da capital brasileira;
3. **Correlação e modelagem preditiva:** avaliação de relações entre poluentes e variáveis temporais, com ajuste de modelos de regressão para medir e prever concentrações atmosféricas.

Os resultados apresentados oferecem um panorama comparativo relevante sobre o comportamento da qualidade do ar em grandes centros urbanos da América Latina.

## 1.1 Estrutura do Relatório

Este documento está organizado da seguinte forma:

- Na Seção 2, apresentamos conceitos fundamentais, padrões climáticos e os poluentes analisados.
- Na Seção 3, detalhamos as fontes de dados e a metodologia utilizada no projeto.
- Na Seção 4, discutimos os principais resultados empíricos obtidos.
- Na Seção 5, interpretamos esses resultados e suas possíveis causas.
- Na Seção 6, sintetizamos as conclusões do relatório.

## 2 Fundamentação Teórica

A qualidade do ar em centros urbanos é determinada pela interação entre fontes de emissão de poluentes, características meteorológicas e condições geográficas que influenciam a dispersão, transporte e remoção de contaminantes atmosféricos. Nesta seção, apresentamos os principais poluentes analisados, os limites estabelecidos pela Organização Mundial da Saúde (OMS) em suas diretrizes de 2021 e os processos físico-atmosféricos que condicionam o comportamento desses elementos na atmosfera.

### 2.1 Poluentes Atmosféricos Analisados

Foram analisados seis poluentes-chave, reconhecidos internacionalmente como indicadores da qualidade do ar: PM<sub>2.5</sub>, PM<sub>10</sub>, NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub> e CO. Cada um apresenta fontes e comportamentos atmosféricos distintos.

#### 2.1.1 Material Particulado Fino (PM<sub>2.5</sub>)

Partículas finas (<2,5 µm) que penetram profundamente nos pulmões e corrente sanguínea. Principais fontes: combustão de combustíveis fósseis, veículos, indústrias e queimadas. Sensível a inversões térmicas e umidade relativa.

#### 2.1.2 Material Particulado Grosso (PM<sub>10</sub>)

Partículas inaláveis (<10 µm), provenientes de poeira do solo, tráfego, indústrias e construção civil. Menos nocivas que PM<sub>2.5</sub>, mas ainda apresentam risco respiratório. Dependem de vento, precipitação e relevo local.

### 2.1.3 Dióxido de Nitrogênio ( $\text{NO}_2$ )

Gás tóxico emitido por veículos e processos de combustão, precursor de partículas secundárias e ozônio troposférico. Concentração associada a tráfego intenso e ventilação atmosférica.

### 2.1.4 Ozônio Troposférico ( $\text{O}_3$ )

Formado fotoquimicamente a partir de  $\text{NO}_x$  e compostos orgânicos voláteis sob radiação solar. Tende a aumentar em climas quentes e secos e regiões com alta emissão veicular ou industrial.

### 2.1.5 Dióxido de Enxofre ( $\text{SO}_2$ )

Emitido pela queima de combustíveis fósseis com enxofre e processos metalúrgicos. Contribui para chuva ácida e partículas sulfatadas, variando conforme perfil energético e qualidade do combustível.

### 2.1.6 Monóxido de Carbono (CO)

Produzido por combustão incompleta de combustíveis em veículos, indústrias e aparelhos domésticos. Reduz o transporte de oxigênio no organismo e se acumula em condições de baixa ventilação urbana.

## 2.2 Diretrizes da OMS (2021)

Em 2021, a Organização Mundial da Saúde atualizou suas Diretrizes Globais de Qualidade do Ar, reforçando limites significativamente mais rigorosos para proteger a saúde pública. A Tabela 1 apresenta os valores-guia relevantes.

Tabela 1: Limites de qualidade do ar segundo a OMS (2021).

<b>Poluente</b>	<b>Período de média</b>	<b>Limite OMS 2021</b>
$\text{PM}_{2.5}$	24 h	$15 \mu\text{g}/\text{m}^3$
$\text{PM}_{2.5}$	Anual	$5 \mu\text{g}/\text{m}^3$
$\text{PM}_{10}$	24 h	$45 \mu\text{g}/\text{m}^3$
$\text{PM}_{10}$	Anual	$15 \mu\text{g}/\text{m}^3$
$\text{NO}_2$	24 h	$25 \mu\text{g}/\text{m}^3$
$\text{NO}_2$	Anual	$10 \mu\text{g}/\text{m}^3$
$\text{O}_3$	8 h (máx. diária)	$100 \mu\text{g}/\text{m}^3$
$\text{SO}_2$	24 h	$40 \mu\text{g}/\text{m}^3$
CO	24 h	$4 \text{ mg}/\text{m}^3$

Esses limites são amplamente utilizados como referência internacional para políticas públicas e avaliações de risco sanitário.

### 2.3 Mecanismos de Dispersão Atmosférica

A dispersão de poluentes depende das características físicas da atmosfera, como estabilidade, turbulência, circulação local e velocidade do vento. A topografia regional desempenha papel crítico: vales estreitos e barreiras montanhosas inibem a ventilação, favorecendo o aprisionamento de poluentes.

Eventos meteorológicos como inversões térmicas — camadas de ar quente que impedem a ascensão do ar frio e poluído — são particularmente relevantes em cidades andinas. A ausência de precipitação, a baixa umidade e as temperaturas amenas reforçam o acúmulo de poluentes, especialmente durante o inverno austral.

## 3 Dados e Metodologia

### 3.1 Base de Dados

Foram utilizados dados horários de qualidade do ar e variáveis meteorológicas para capitais latino-americanas, provenientes do conjunto *Latin America Weather and Air Quality Data* (Kaggle), derivado da API *Open-Meteo*. O dataset contempla PM<sub>2.5</sub>, PM<sub>10</sub>, NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub>, CO, temperatura, umidade relativa, precipitação e velocidade do vento.

Tabela 2: Variáveis empregadas na análise.

Variável	Descrição	Unidade
PM <sub>2.5</sub>	Material particulado fino	$\mu\text{g}/\text{m}^3$
PM <sub>10</sub>	Material particulado inalável	$\mu\text{g}/\text{m}^3$
NO <sub>2</sub>	Dióxido de nitrogênio	ppb
SO <sub>2</sub>	Dióxido de enxofre	ppb
O <sub>3</sub>	Ozônio troposférico	ppb
CO	Monóxido de carbono	ppm
Temperatura	Temperatura do ar	°C
Umidade Relativa	Conteúdo de vapor d'água	%
Precipitação	Volume de chuva	mm
Velocidade do Vento	Intensidade do vento	m/s

### 3.1.1 Padronização e Pré-Processamento

Inicialmente, realizou-se a harmonização das estruturas tabulares, incluindo a padronização dos nomes das variáveis, conversão de datas para o formato `datetime`, e verificação da consistência das unidades. Procedeu-se à inspeção de valores ausentes, seguidos de remoção de casos inválidos ou interpolação temporal quando estatisticamente justificável.

## 3.2 Análise Exploratória de Dados (EDA)

A análise exploratória buscou caracterizar o comportamento temporal e estatístico das concentrações de poluentes. Foram empregados:

- estatísticas descritivas (média, mediana, quartis e variância);
- gráficos de violino, boxplots e calendários de calor para identificação de sazonalidades e outliers;
- quantificação da proporção de dias acima dos limites recomendados pela OMS.

Esta etapa permitiu identificar padrões estruturais, tais como picos sazonais recorrentes.

## 3.3 Testes de Hipóteses

### 3.3.1 Teste F de Fisher para Igualdade de Variâncias

Antes de comparar as médias de concentração de poluentes entre Brasília e as demais capitais latino-americanas, verificou-se se as variâncias das duas amostras podiam ser assumidas como iguais. Para isso, utilizou-se o *Teste F de Fisher*, adequado para comparar a variabilidade de duas populações assumindo normalidade.

Sejam  $s_B^2$  e  $s_O^2$  as variâncias amostrais de Brasília e das outras capitais, respectivamente. A estatística de teste é dada por:

$$F = \frac{s_B^2}{s_O^2},$$

a qual, sob a hipótese nula, segue uma distribuição  $F$  com  $n_B - 1$  e  $n_O - 1$  graus de liberdade.

As hipóteses consideradas foram:

$$H_0 : \sigma_B^2 = \sigma_O^2, \quad H_1 : \sigma_B^2 \neq \sigma_O^2.$$

O teste foi conduzido com nível de significância  $\alpha = 0,05$ , servindo como etapa preliminar para determinar se a comparação entre as médias deveria ser feita por meio do teste t clássico (variâncias iguais) ou pelo teste de Welch (variâncias desiguais).

### 3.3.2 Teste de Welch

Para comparar as médias de concentração de poluentes entre Brasília e as demais capitais latino-americanas, utilizou-se o *teste de Welch*, apropriado para amostras com variâncias desconhecidas e desiguais.

Sejam  $\bar{X}_B$  e  $\bar{X}_O$  as médias amostrais de Brasília e das outras capitais, com variâncias  $s_B^2$  e  $s_O^2$  e tamanhos  $n_B$  e  $n_O$ . A estatística de teste é:

$$t = \frac{\bar{X}_B - \bar{X}_O}{\sqrt{\frac{s_B^2}{n_B} + \frac{s_O^2}{n_O}}}$$

com graus de liberdade aproximados por:

$$df \approx \frac{\left(\frac{s_B^2}{n_B} + \frac{s_O^2}{n_O}\right)^2}{\frac{(s_B^2/n_B)^2}{n_B-1} + \frac{(s_O^2/n_O)^2}{n_O-1}}.$$

A hipótese nula ( $H_0$ ) assume igualdade das médias:

$$H_0 : \mu_B = \mu_O, \quad H_1 : \mu_B \neq \mu_O,$$

e a decisão estatística foi tomada com nível de significância  $\alpha = 0,05$ .

## 3.4 Correlação e Modelagem

A relação linear entre poluentes e variáveis meteorológicas foi quantificada pelo coeficiente de Pearson.

### 3.4.1 Modelo Preditivo para PM<sub>2.5</sub>

Adicionalmente, foi ajustado modelo de regressão linear múltipla para estimar PM<sub>2.5</sub> como variável resposta. Considerando as features selecionadas — temperatura ( $T$ ), umidade relativa ( $U$ ), velocidade do vento ( $V$ ), PM<sub>2.5</sub> do dia anterior (PM<sub>2.5</sub><sup>lag</sup>), mês ( $M$ ), dia da semana ( $D$ ) e cidade ( $C$ ) — o modelo pode ser expresso como:

$$\widehat{\text{PM}_{2.5}} = \beta_0 + \beta_1 T + \beta_2 U + \beta_3 V + \beta_4 \text{PM}_{2.5}^{\text{lag}} + \sum_i \beta_{5,i} \mathbf{1}_{M=i} + \sum_j \beta_{6,j} \mathbf{1}_{D=j} + \sum_k \beta_{7,k} \mathbf{1}_{C=k} + \varepsilon,$$

onde 1. são variáveis indicadoras (*one-hot*) para capturar efeitos categóricos, e  $\varepsilon$  é o termo de erro.

A inclusão do lag temporal e das variáveis categóricas permitiu ao modelo capturar dependências temporais, padrões sazonais e efeitos específicos de cada cidade, enquanto as variáveis contínuas representam o efeito direto das condições meteorológicas sobre a concentração de PM<sub>2.5</sub>.

O modelo foi avaliado via  $R^2$  e erro quadrático médio (RMSE).

### 3.4.2 Modelo Explicativo para PM<sub>2.5</sub>

A Figura 1 sintetiza o procedimento metodológico adotado neste projeto. O fluxo inicia no pré-processamento e divisão dos dados, segue pela aplicação das três abordagens de seleção de hiperparâmetros (LassoCV, Grid Search e Optuna) e culmina na avaliação comparativa dos modelos por meio de métricas padronizadas ( $R^2$ , VIF e seleção de variáveis).

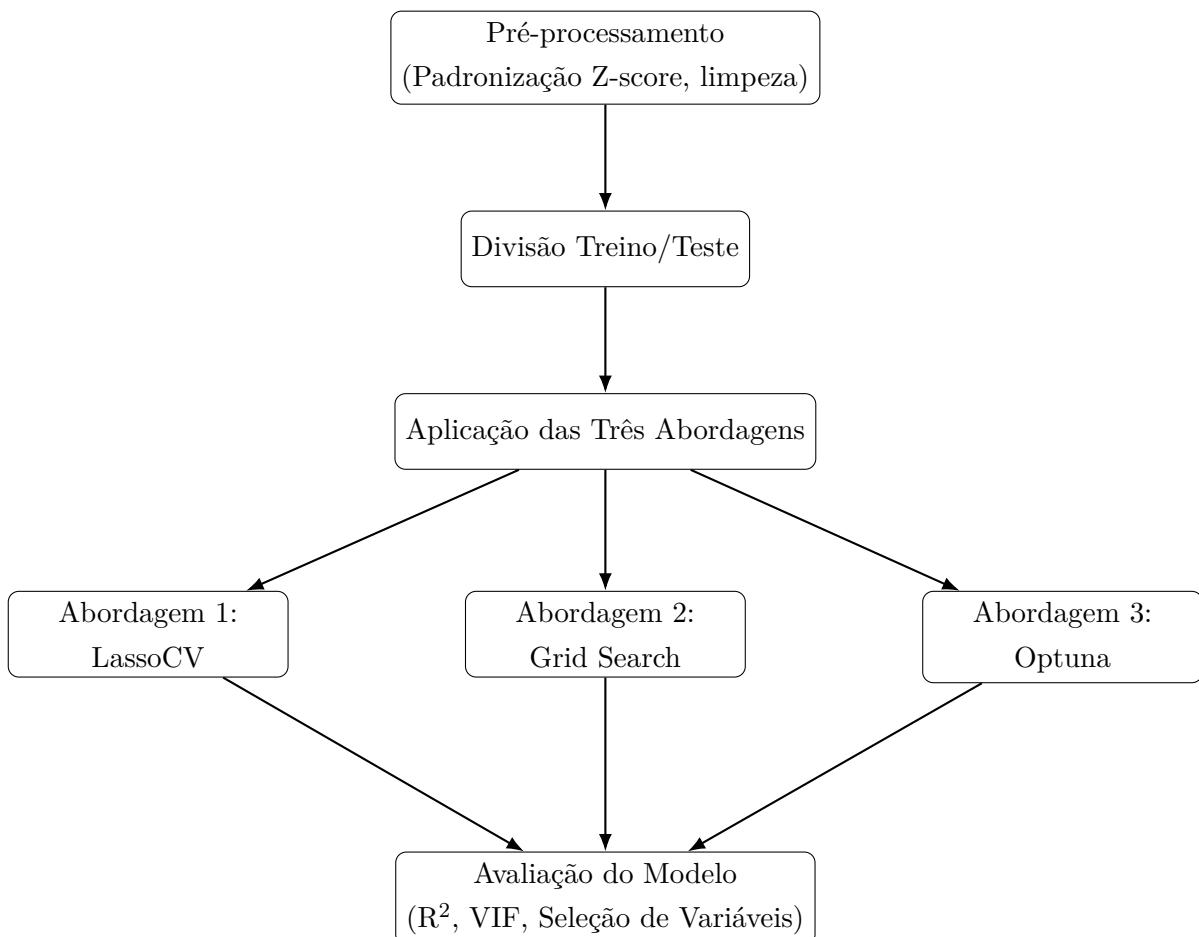


Figura 1: Fluxograma da metodologia de modelagem e seleção de hiperparâmetros

### 3.5 Ferramentas Computacionais

As análises foram realizadas em Python, utilizando:

- **Pandas** para manipulação de dados;
- **NumPy** para operações vetorializadas;
- **Matplotlib** para visualização;
- **SciPy** para testes estatísticos;
- **scikit-learn** para regressões.

### 3.6 Limitações da Análise

Apesar dos resultados contribuírem para uma compreensão comparativa da qualidade do ar nas capitais latino-americanas, algumas limitações devem ser consideradas:

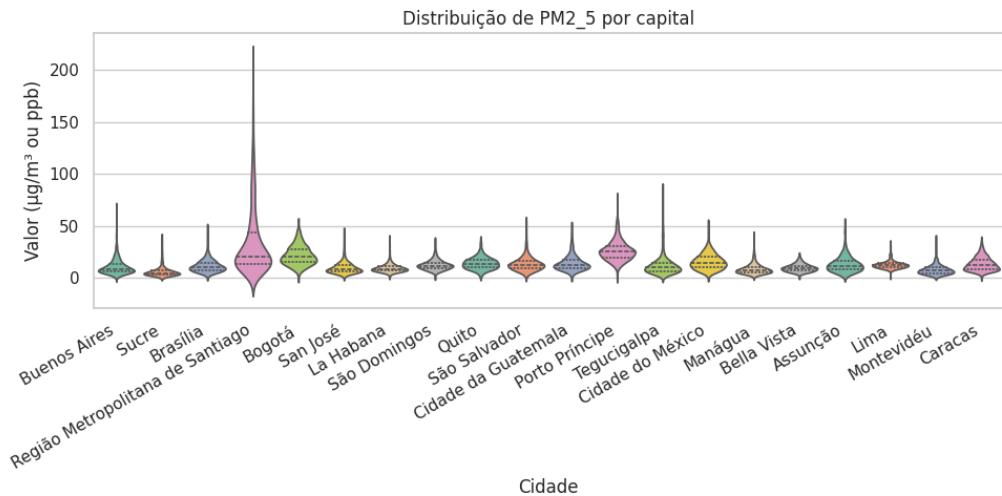
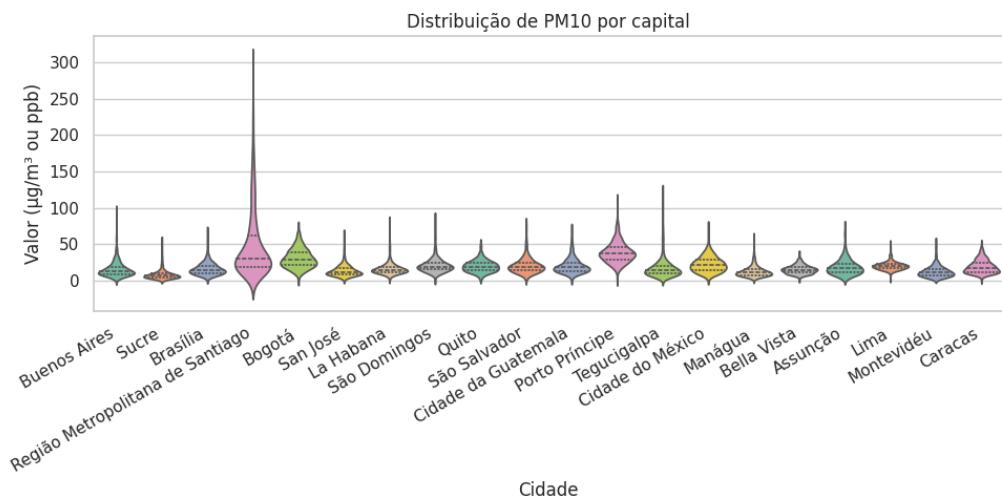
- **Intervalo temporal reduzido:** os dados cobrem poucos anos, dificultando conclusões de longo prazo.
- **Foco apenas nas capitais:** resultados não podem ser extrapolados para todo o território nacional.
- **Ausência de especialistas:** as interpretações são fundamentadas, mas não substituem análises meteorológicas ou geográficas profissionais.

## 4 Resultados

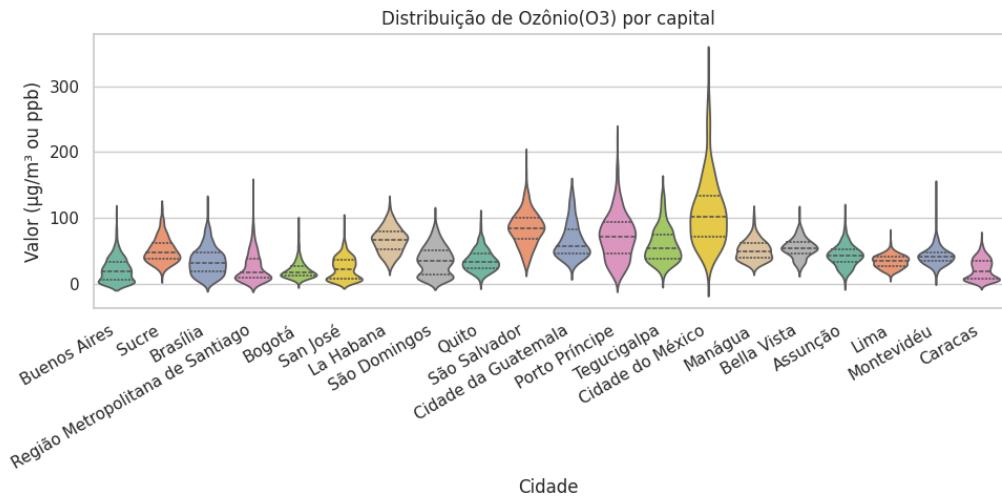
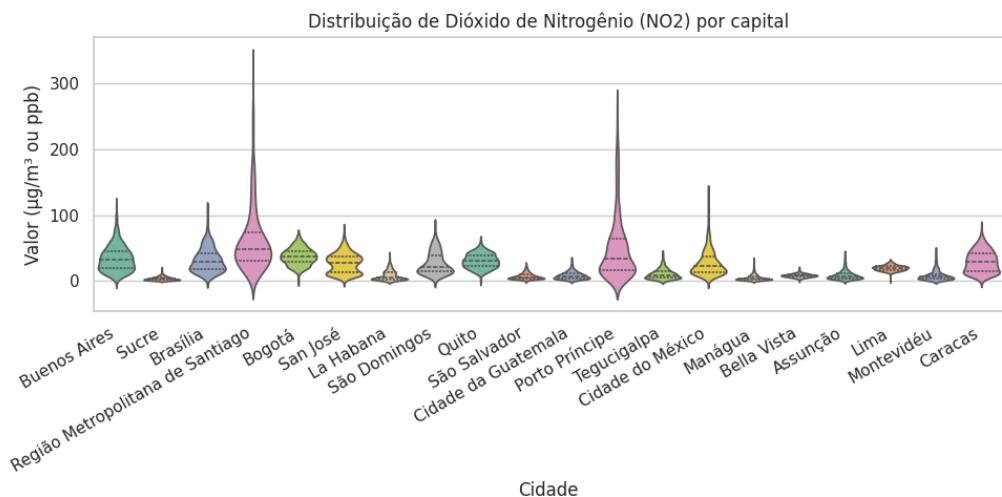
### 4.1 Estatísticas Descritivas e Distribuições

A caracterização estatística das concentrações de poluentes nas capitais latino-americanas requer a análise de suas distribuições empíricas, de forma a identificar padrões estruturais, como mediana elevada, assimetria acentuada ou variabilidade atípica. Para esse fim, foram construídos gráficos de violino para cada indicador atmosférico presente na base de dados, permitindo visualizar simultaneamente a densidade das observações, a dispersão e os principais quantis.

Com o objetivo de manter a objetividade da apresentação, destacam-se nesta subseção apenas os poluentes cujas distribuições apresentaram maior relevância analítica; as demais visualizações encontram-se no Apêndice A. As Figuras 2 e 3 ilustram as distribuições de material particulado fino ( $PM_{2.5}$ ) e grosso ( $PM_{10}$ ) nas capitais analisadas.

Figura 2: Distribuição de PM<sub>2.5</sub> nas capitais latino-americanas.Figura 3: Distribuição de PM<sub>10</sub> nas capitais latino-americanas.

A Região Metropolitana de Santiago (Chile) apresenta destaque marcante, com dispersão elevada e caudas longas nas distribuições de PM<sub>2.5</sub> e PM<sub>10</sub>. O comportamento observado (mediana: 30,4 ng/m<sup>3</sup>; média: 49,69; Q1: 19,4; Q3: 62,4; máximo: 292,4) é consistente com episódios recorrentes de acúmulo de poluentes associados à configuração geográfica da cidade — situada em um vale cercado pela Cordilheira dos Andes, condição que reduz a ventilação horizontal e favorece inversões térmicas. A alta densidade urbana, o tráfego intenso e as emissões industriais e residenciais reforçam esse padrão de concentração.

Figura 4: Distribuição de O<sub>3</sub> nas capitais latino-americanas.Figura 5: Distribuições de NO<sub>2</sub> nas capitais latino-americanas.

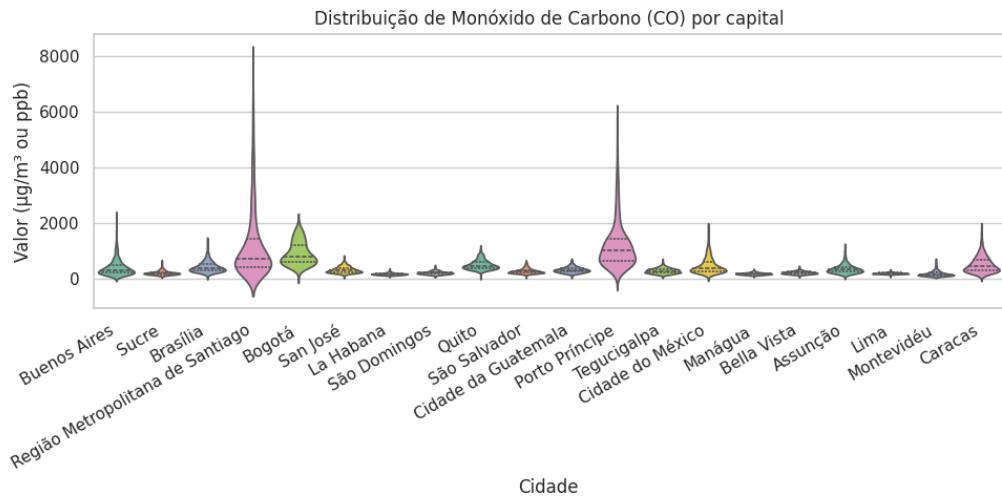


Figura 6: Distribuições de CO nas capitais latino-americanas.

Para os gases O<sub>3</sub>, NO<sub>2</sub> e CO, observa-se um comportamento destoante em Porto Príncipe (Haiti), cuja distribuição evidencia caudas longas e valores persistentemente elevados. A interpretação mais plausível envolve a predominância de combustíveis de baixa qualidade, uma frota veicular envelhecida e pouco regulada, além da limitação de políticas ambientais e infraestrutura de controle de emissões, resultando em níveis elevados de gases tóxicos e intensificação da formação fotoquímica de ozônio.

Outros gráficos referentes a poluentes não exibidos nesta subseção estão disponíveis para consulta no Apêndice B.1 Gráficos de Violino.

#### 4.1.1 Avaliação de Capitais Via Métricas de Qualidade do Ar

Após analisar como, temporalmente, a distribuição dos gases e dos materiais particulares se comportam nas capitais latino-americanas, foi de interesse do grupo obter limiares que permitam a metrificação da situação das capitais no que diz respeito à concentração de indicadores de risco à saúde humana. Para isto, utilizou-se, como referência, os limites de alerta mais recentemente atualizados pela Organização Mundial de Saúde (OMS) em 2021, conforme pode-se consultar na subseção 2.2.

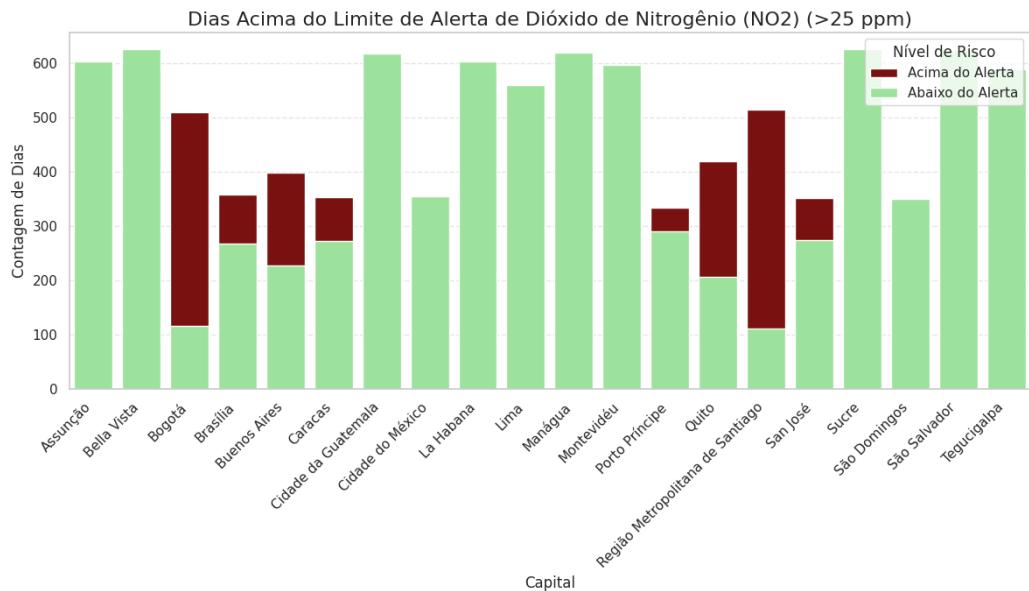


Figura 7: Proporção de dias acima do limite estabelecido pela OMS para Dióxido de Nitrogênio.

A avaliação dos indicadores de qualidade do ar revela três casos particularmente relevantes. Para *Dióxido de Nitrogênio (NO<sub>2</sub>)*, Santiago se destaca com mais de 70% dos dias acima do limite da OMS, resultado consistente com sua topografia em bacia cercada pelos Andes e frequentes inversões térmicas. Bogotá e Quito exibem padrão semelhante, influenciado por altitudes elevadas, vales restritivos e tráfego intenso. Mesmo cidades planas, como Buenos Aires e Brasília, registram episódios pontuais de NO<sub>2</sub> elevado, geralmente associados a tráfego intenso e condições sazonais desfavoráveis.

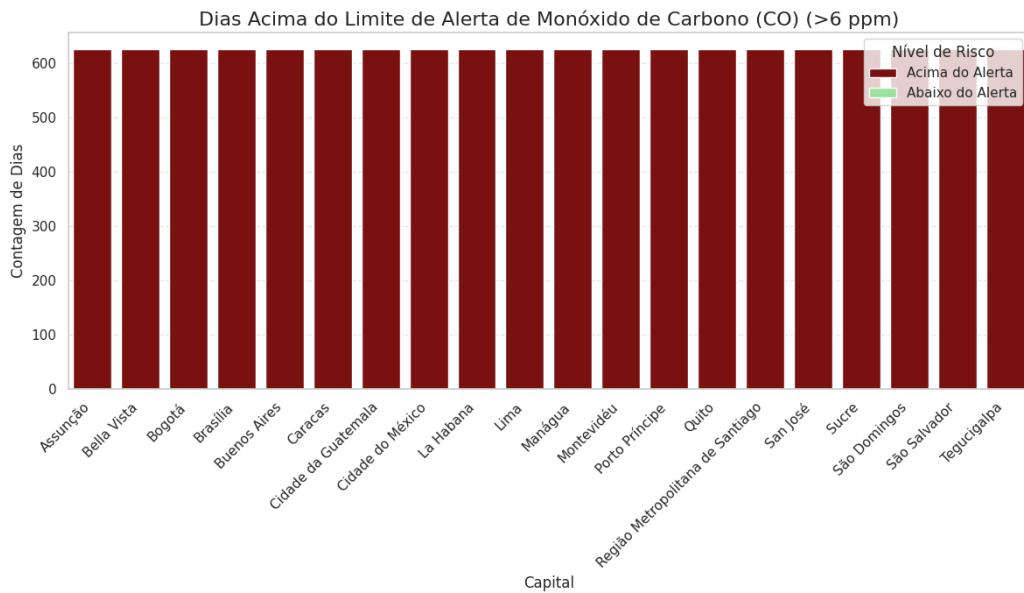


Figura 8: Proporção de dias acima do limite estabelecido pela OMS para Monóxido de Carbono.

No caso do *Monóxido de Carbono (CO)*, todas as capitais latino-americanas apresentaram concentrações acima dos limites em 100% dos dias analisados. Esse comportamento homogêneo decorre da forte dependência regional de combustíveis fósseis no transporte, associada a condições meteorológicas que dificultam a dispersão em diferentes centros urbanos.

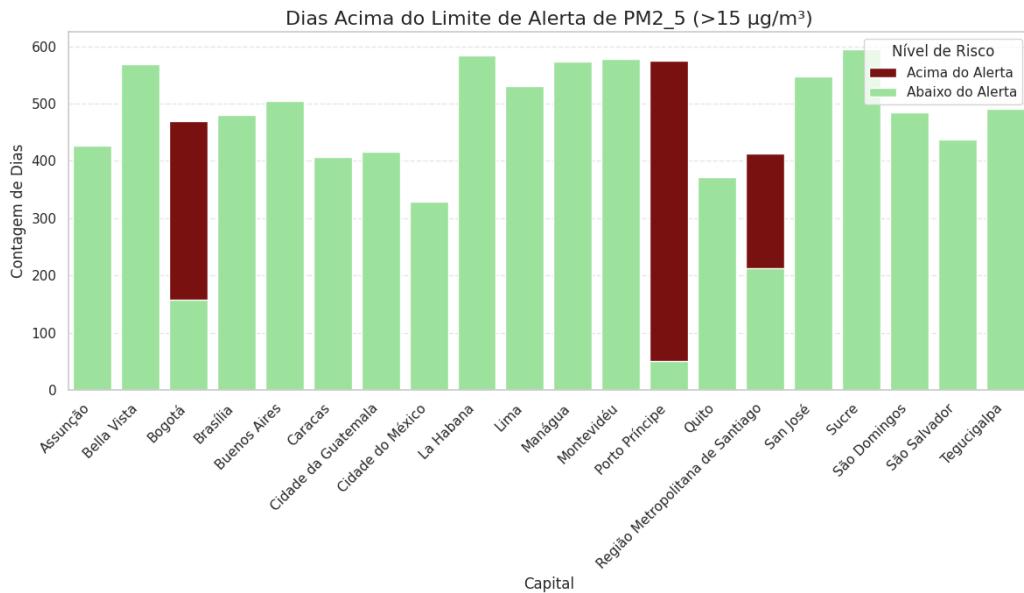


Figura 9: Proporção de dias acima do limite estabelecido pela OMS para PM<sub>2.5</sub>.

Por fim, para *Material Particulado Fino (PM<sub>2.5</sub>)*, Porto Príncipe apresentou o cenário mais crítico, com mais de 90% dos dias acima do limite da OMS, impulsio-

nado por poluição veicular, industrial e emissões de poeira urbana. Bogotá e Santiago também excedem o padrão em proporções relevantes (60% e 40% dos dias, respectivamente), influenciadas por topografia e sazonalidade. Brasília, em contraste, não registrou nenhum dia acima do limite, configurando-se como um ponto fora do padrão observado nas demais capitais.

Os demais gráficos gerados para esta subseção podem ser consultados no Apêndice B.2 Gráficos de Limites.

## 4.2 Padrões Sazonais e Mensais

A caracterização da sazonalidade dos poluentes atmosféricos foi conduzida inicialmente por meio de gráficos de boxplot mensais, que permitem identificar mudanças estruturais na mediana e na variabilidade intra-anual. Entre os indicadores avaliados, três poluentes se destacam por apresentar padrões robustos e informativos — PM<sub>2.5</sub>, PM<sub>10</sub> e O<sub>3</sub> — sendo, portanto, discutidos diretamente nesta subseção. Os resultados referentes a CO, NO<sub>2</sub> e SO<sub>2</sub>, embora tecnicamente consistentes, não acrescentam novos elementos interpretativos e são apresentados integralmente no Apêndice B.3 Boxplots.

### Materiais Particulados (PM<sub>2.5</sub> e PM<sub>10</sub>)

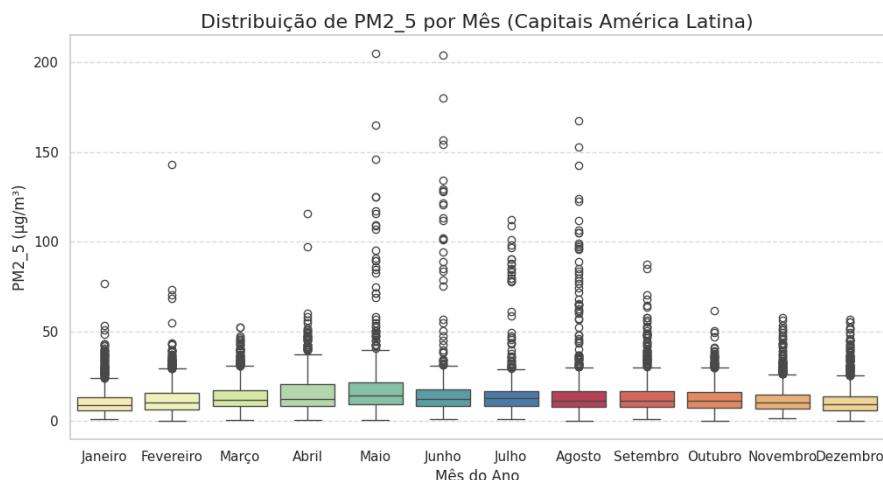


Figura 10: Distribuição de PM<sub>2.5</sub> nas capitais latino-americanas.

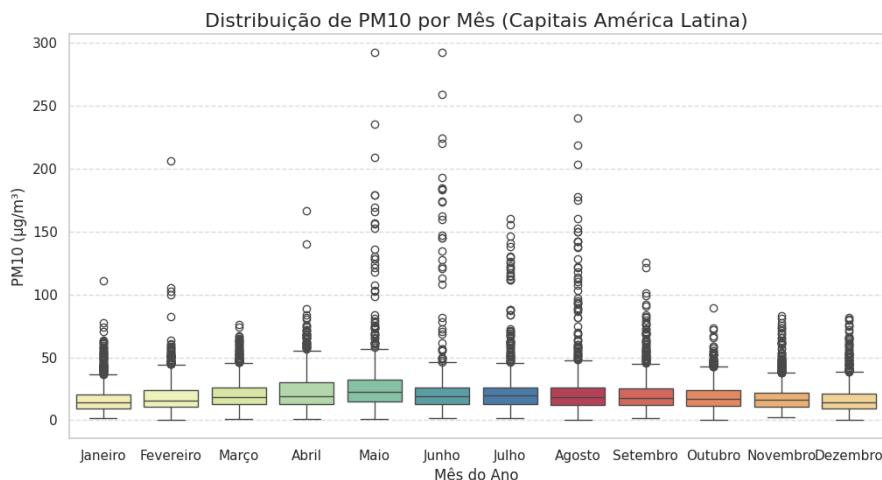


Figura 11: Distribuição de PM<sub>10</sub> nas capitais latino-americanas.

As Figuras 10 e 11 evidenciam um padrão sazonal claro nos materiais particulados, com elevação consistente das medianas em maio, seguida de declínio acentuado a partir de junho. Essa dinâmica é compatível com o regime climático da maior parte da América Latina: o avanço do período úmido intensifica a remoção de partículas por deposição úmida, reduzindo simultaneamente os níveis medianos e a dispersão das concentrações.

### Ozônio Troposférico (O<sub>3</sub>)

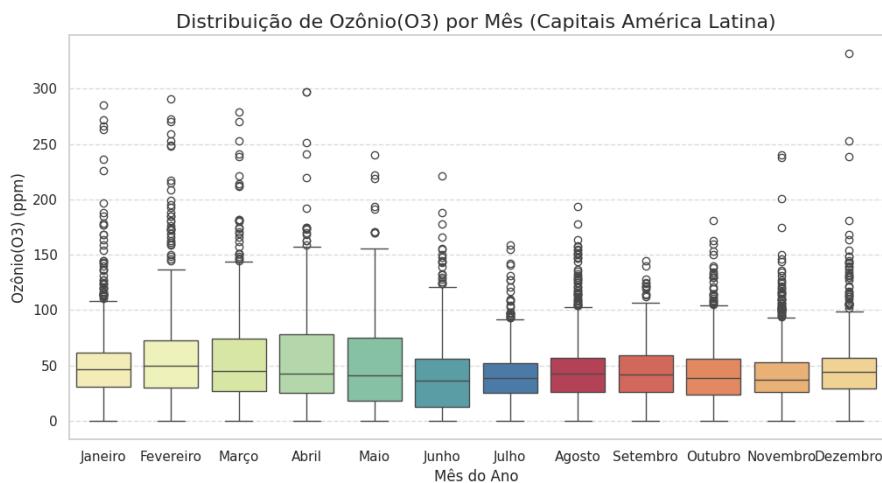


Figura 12: Distribuição de O<sub>3</sub> nas capitais latino-americanas.

O comportamento do O<sub>3</sub>, mostrado na Figura 12, apresenta amplitudes maiores que as observadas para os particulados, refletindo seu caráter fotoquímico. Meses com maior radiação solar tendem a registrar maior dispersão e picos de concentração, reforçando

o efeito combinado de NO<sub>x</sub>, compostos orgânicos voláteis e luz solar na formação do ozônio troposférico.

## Síntese dos Padrões Observados

De modo geral, os três poluentes revelam medianas mais baixas no segundo semestre, período meteorologicamente mais úmido e favorável à dispersão e deposição dos contaminantes. Esse conjunto de padrões, capturado pelos boxplots, orienta a etapa seguinte da análise, voltada para a resolução temporal diária dos processos sazonais.

### 4.2.1 Calendários de Calor (Heatmaps)

Para aprofundar a interpretação da sazonalidade identificada nos boxplots, foram construídos calendários de calor para PM<sub>2.5</sub>, PM<sub>10</sub> e O<sub>3</sub>, utilizando o ano de 2023 — o único com série completa; os demais podem ser consultados no Apêndice B.4 Calendários de Calor. Esses gráficos permitem observar a persistência dos episódios críticos, transições entre regimes atmosféricos e a organização diária dos períodos de maior acúmulo de poluentes.

As Figuras 13–18 apresentam os heatmaps selecionados. Nos materiais particulados, observa-se nas capitais um aumento concentrado em maio, enquanto em Brasília o pico migra para setembro, em consonância com a estação seca do Centro-Oeste. O ozônio, por sua vez, exibe intensificação entre agosto e novembro, alinhada ao período de maior radiação solar.

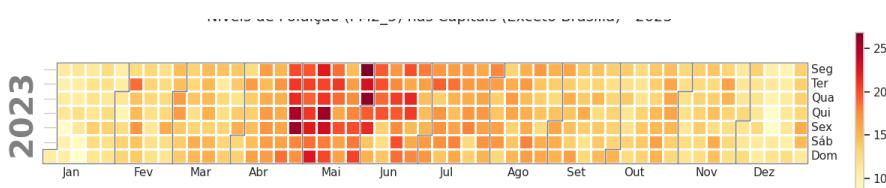


Figura 13: Calendário de calor de PM<sub>2.5</sub> – Capitais latino-americanas (2023).

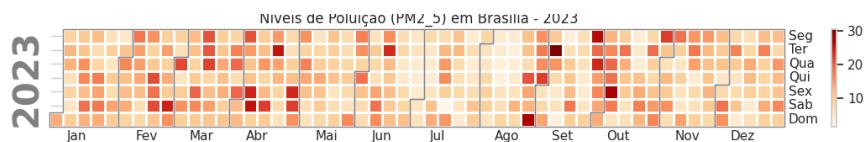


Figura 14: Calendário de calor de PM<sub>2.5</sub> – Brasília (2023).

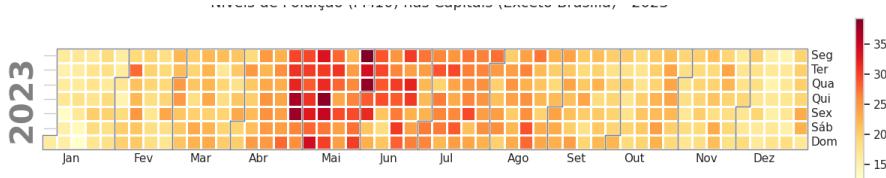


Figura 15: Calendário de calor de PM<sub>10</sub> – Capitais latino-americanas (2023).

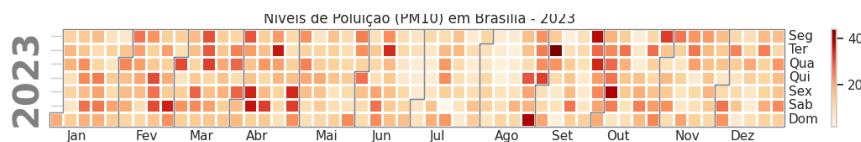


Figura 16: Calendário de calor de PM<sub>10</sub> – Brasília (2023).

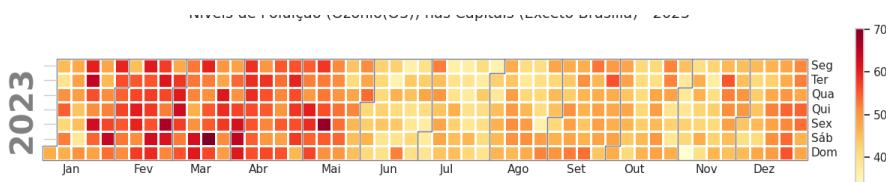


Figura 17: Calendário de calor de O<sub>3</sub> – Capitais latino-americanas (2023).

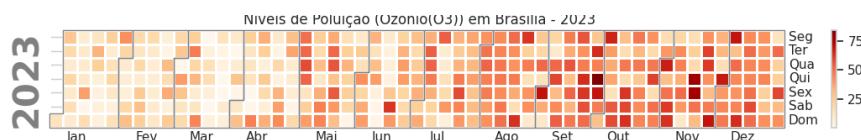


Figura 18: Calendário de calor de O<sub>3</sub> – Brasília (2023).

### 4.3 Comparação Entre Capitais

Nesta etapa, avaliamos se Brasília apresenta níveis médios de poluição atmosférica significativamente distintos daqueles observados no conjunto das demais capitais latino-americanas. Inicialmente, procedeu-se com uma análise exploratória por meio de Intervalos de Confiança (IC) de 95% para as médias amostrais dos poluentes, computados separadamente para Brasília e para o grupo das demais capitais. Essa abordagem permite verificar, de forma visual, se as estimativas pontuais da capital federal se situam dentro ou fora das regiões de maior verossimilhança para o restante da amostra.

Entre todos os indicadores avaliados, o resultado mais informativo refere-se ao Monóxido de Carbono (CO), apresentado na Figura 19.

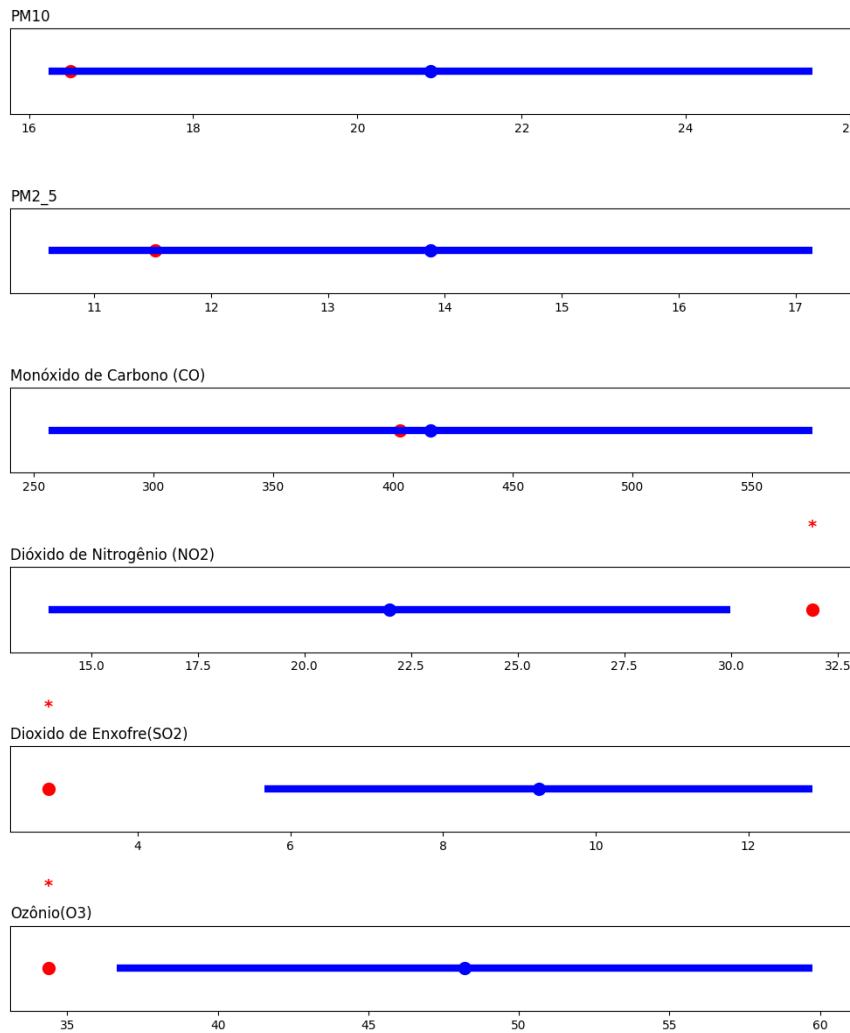


Figura 19: Em terceiro, Intervalo de Confiança (95%) para a média de CO: Brasília vs. demais capitais latino-americanas.

Observa-se que o IC para as demais capitais contém a estimativa pontual de Brasília, além de ambas as médias se situarem em posições próximas. Esse alinhamento antecipou o principal resultado inferencial obtido na subseção seguinte: entre todos os poluentes avaliados, apenas o CO não apresentou evidências estatísticas suficientes para rejeitar a igualdade entre as médias populacionais.

#### 4.4 Resultados dos Testes de Hipótese

Para formalizar as conclusões da análise exploratória, realizamos testes de hipótese para todos os poluentes, comparando Brasília ao conjunto das demais capitais. Cada teste seguiu dois passos: (i) avaliação da igualdade das variâncias amostrais e (ii) aplicação do teste t adequado ao cenário detectado.

## Teste de Igualdade de Variâncias

A comparação inicial entre variâncias foi realizada via Teste F, utilizando nível de significância  $\alpha = 0,05$ . Para todos os poluentes, incluindo o CO, identificou-se heterocedasticidade. Estes resultados podem ser consultados no Apêndice 7. Para o CO, especificamente, temos:

Variância de Brasília ( $s_1^2$ )	32,295.21
Variância das demais capitais ( $s_2^2$ )	258,035.90
Estatística F observada	0.13
Região de aceitação	[0.89, 1.12]
Conclusão	Rejeita-se $H_0$ (variâncias diferentes)

Identificada a heterocedasticidade, procedeu-se ao Teste t de Welch para avaliar diferenças entre médias.

## Teste t de Welch

As hipóteses testadas para cada poluente foram:

$$H_0 : \mu_{\text{Brasília}} = \mu_{\text{capitais}} \quad \text{vs.} \quad H_1 : \mu_{\text{Brasília}} \neq \mu_{\text{capitais}}.$$

A Tabela 3 resume os resultados obtidos.

Tabela 3: Resultados do Teste t de Welch: Brasília vs. demais capitais.

Poluente	$\bar{x}_B$	$\bar{x}_C$	$t$	$p$ -valor
PM <sub>10</sub>	16.50	20.89	-11.67	$1.99 \times 10^{-29}$
PM <sub>2.5</sub>	11.52	13.88	-9.00	$1.26 \times 10^{-18}$
CO	402.91	415.54	-1.47	$1.41 \times 10^{-1}$
NO <sub>2</sub>	31.91	21.99	13.26	$2.89 \times 10^{-36}$
SO <sub>2</sub>	2.83	9.25	-46.40	$< 10^{-15}$
O <sub>3</sub>	34.37	48.19	-15.38	$6.58 \times 10^{-47}$

Entre todos os poluentes avaliados, o CO é o único cujo valor- $p$  ( $p = 0,1405$ ) excede o nível de significância adotado, impedindo a rejeição de  $H_0$ . Dessa forma, \*\*não se identificou evidência estatística de que a concentração média de CO em Brasília difira da média observada nas demais capitais latino-americanas\*\*.

Esse resultado converge plenamente com a interpretação visual fornecida pelos Intervalos de Confiança e reforça a consistência entre os diferentes métodos empregados. Além disso, está alinhado às observações realizadas na Seção ??, onde se discute que todas as capitais, incluindo Brasília, apresentam níveis persistentemente elevados de

CO ao longo do período analisado, frequentemente acima das diretrizes de alerta da OMS.

## 4.5 Correlação Entre Variáveis

### 4.5.1 Correlação Entre Gases e Variáveis Meteorológicas

Um dos objetivos centrais desta pesquisa é avaliar a viabilidade de modelos preditivos capazes de estimar os níveis de PM<sub>2.5</sub> nas capitais latino-americanas a partir de informações meteorológicas e de séries históricas de poluentes. Para isso, investigamos a existência de relações lineares entre variáveis atmosféricas (temperatura média, precipitação acumulada, velocidade do vento e evapotranspiração) e os diferentes poluentes monitorados. Foram construídos gráficos de dispersão para todos os pares possíveis entre poluentes e variáveis meteorológicas, complementados pelo cálculo do coeficiente de correlação linear de Pearson.

Os resultados revelaram um padrão amplamente homogêneo: os coeficientes de correlação raramente ultrapassam  $r = 0.3$ , indicando associações fracas entre os conjuntos de variáveis. Os gráficos de dispersão corroboram esse diagnóstico, exibindo nuvens de pontos sem estrutura que sugira dependência linear relevante. Em termos práticos, isso implica que nenhuma variável climática, isoladamente, detém capacidade preditiva suficiente para modelos de Regressão Linear Simples.

Apesar da baixa magnitude dos coeficientes, observou-se que os testes de significância associados às correlações produziram, em quase todos os casos, valores de  $p$  iguais a 0.0. Tal resultado não reflete força de relação, mas sim o grande tamanho amostral da base, que confere elevado poder estatístico capaz de detectar até efeitos minúsculos com alta confiança. Assim, embora estatisticamente diferentes de zero, as correlações são fracas demais para terem utilidade prática em previsão.

A principal exceção encontrada corresponde à relação entre *Soma\_precipitacao* e o dióxido de nitrogênio (NO<sub>2</sub>). Neste caso, o valor de  $p = 0.0697 > 0.05$  indica ausência de significância estatística, sugerindo que a correlação observada representa apenas ruído amostral.

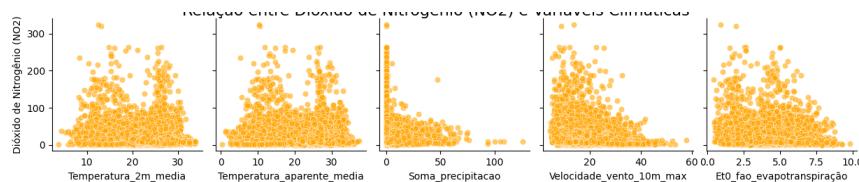


Figura 20: Gráficos de dispersão entre Dióxido de Nitrogênio (NO<sub>2</sub>) e variáveis climáticas.

Em síntese, os resultados demonstram que, mesmo com elevada significância es-

tatística, as associações lineares entre clima e poluentes são estruturalmente fracas ( $r \approx 0$ ). Consequentemente, a previsão de poluentes atmosféricos exige abordagens mais robustas, como modelos de Regressão Linear Múltipla, capazes de explorar combinações de variáveis em vez de dependências simples e isoladas.

Os demais gráficos de dispersão estão disponíveis para consultar no Apêndice B.5 Relação Entre Variáveis Climáticas e Meteorológicas.

#### 4.5.2 Correlação Entre os Gases

Diante da baixa correlação entre variáveis meteorológicas e poluentes, investigamos a possibilidade de relações significativas entre os próprios gases monitorados. Foram produzidos gráficos de dispersão para todos os pares de poluentes, permitindo uma avaliação visual inicial das possíveis relações.

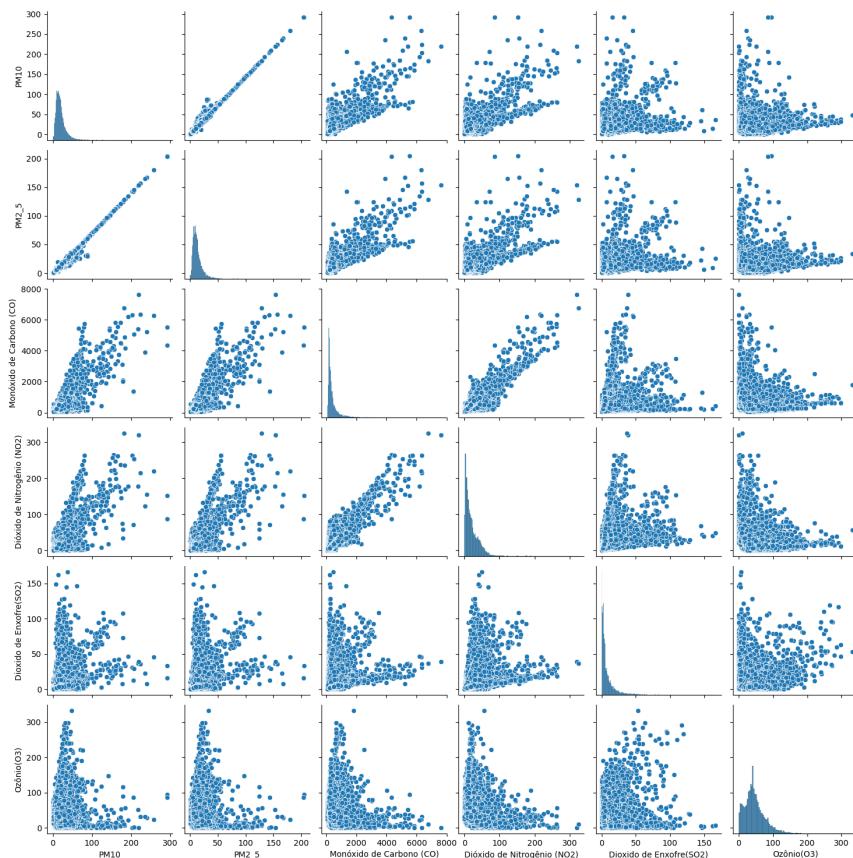


Figura 21: Gráficos de dispersão de relação entre gases.

Em seguida, foram calculados os coeficientes de correlação de Pearson e construída uma matriz de correlação, possibilitando quantificar formalmente as associações detectadas.

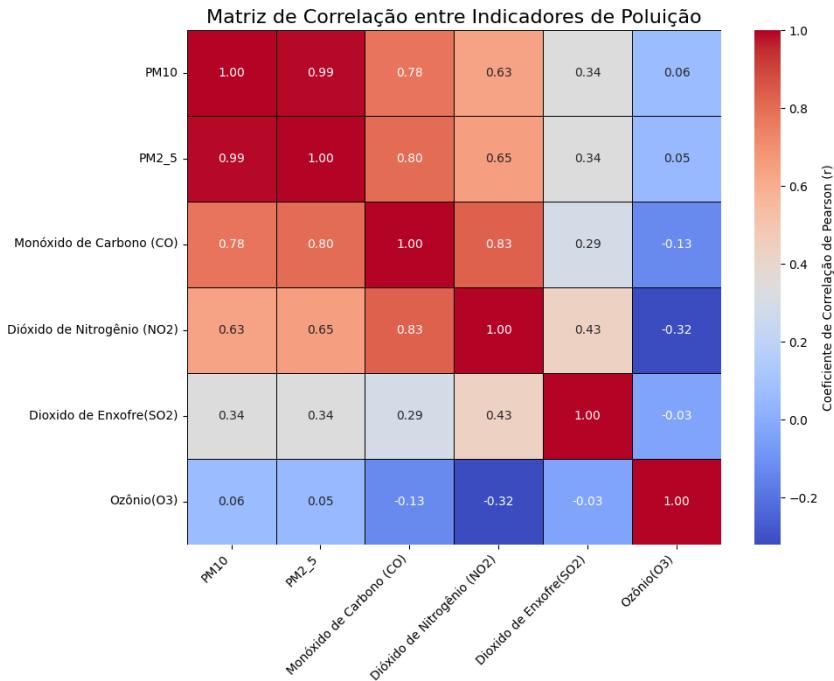


Figura 22: Matriz de Correlação entre gases.

Os resultados revelam a existência de correlações lineares substanciais entre diversos poluentes, em contraste com o cenário clima-poluente. Essa estrutura indica que os gases não variam de forma independente, o que provavelmente reflete fontes emissoras comuns (como frota veicular, queimadas ou processos industriais), além de possíveis interações químicas ou condições atmosféricas que afetam simultaneamente múltiplos poluentes.

Essas conclusões fornecem um insight fundamental: embora variáveis climáticas isoladas tenham baixo poder explicativo, a estrutura conjunta dos poluentes contém informações essenciais para a modelagem preditiva. A partir desse diagnóstico, avançaremos para a construção de modelos regressivos capazes de estimar o PM<sub>2.5</sub> considerando tanto variáveis meteorológicas quanto as dependências estruturais entre os gases.

## 4.6 Modelos Preditivos

Nesta parte, buscamos avaliar se modelos de regressão conseguem prever os níveis diários de PM<sub>2.5</sub> em cidades latino-americanas a partir de dados meteorológicos e históricos. As correlações individuais entre clima e poluição foram fracas, indicando que apenas modelos multivariados seriam capazes de capturar padrões úteis.

#### 4.6.1 Preparação das Features

Foram incluídas variáveis com relevância conhecida para dinâmica de poluição: defasagem temporal do PM2.5, variáveis sazonais (mês e dia da semana), medidas meteorológicas e identificação da cidade. O PM2.5 diário foi definido como variável-alvo.

#### 4.6.2 Estrutura do Modelo

Os dados foram divididos em treino e teste, e um *pipeline* integrou padronização, codificação categórica e Regressão Linear Múltipla. O desempenho foi avaliado por MSE e  $R^2$ , com intervalos de confiança obtidos via *bootstrap*.

#### 4.6.3 Resultados

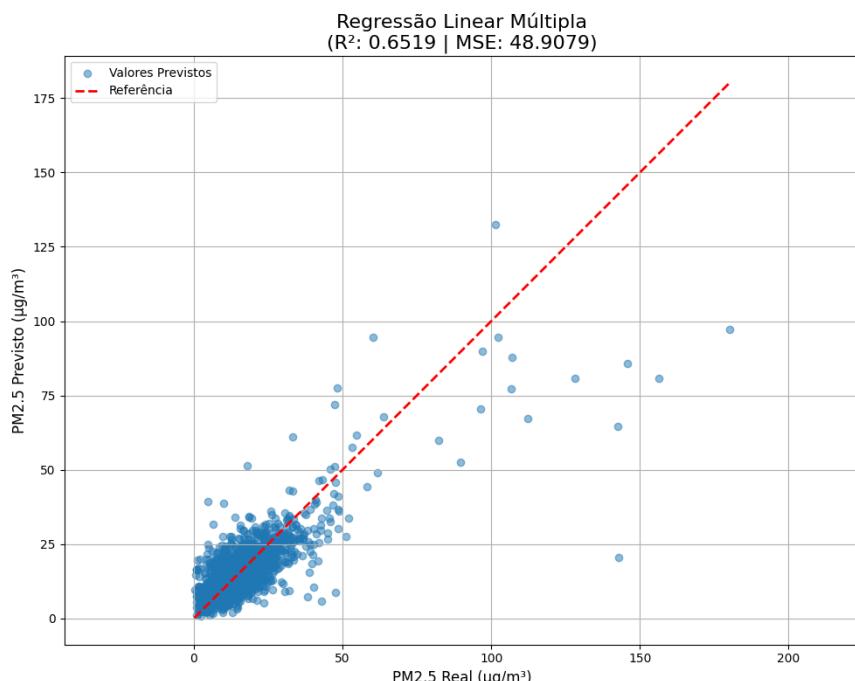


Figura 23: Valores reais vs. preditos.

O modelo linear, ilustrado pelo gráfico acima, apresentou desempenho moderado e consistente, mostrando alinhamento geral com a linha  $y = x$ , mas com dispersão considerável, especialmente para valores altos de PM2.5.

- **R<sup>2</sup>**: 0.6519 no conjunto de teste, indicando explicação de 65% da variabilidade. O intervalo de 95% foi [0.5790, 0.7209].
- **MSE**: 48.9079, equivalente a um RMSE de cerca de  $7 \mu\text{g}/\text{m}^3$ . O intervalo de 95% variou de [33.6249, 64.3762].

#### 4.6.4 Interpretação

Dois padrões se destacam:

- **Subestimação dos picos:** o modelo tende a prever valores abaixo dos observados em episódios de alta poluição.
- **Sinais de não-linearidade:** a dispersão e os erros maiores sugerem que relações não-lineares são relevantes, indicando potencial ganho com modelos mais flexíveis.

#### 4.6.5 Importância das Variáveis por Cidade

Modelos ajustados separadamente para cada cidade, utilizando apenas variáveis meteorológicas, revelaram grande variação regional na importância dos preditores. Coeficientes padronizados permitiram comparações diretas e mostraram que temperatura, precipitação, vento e evapotranspiração influenciam o PM2.5 de forma heterogênea entre localidades, refletindo diferenças climáticas e fontes locais de emissão. Abaixo estão alguns exemplos de elevada variação dos coeficientes, a depender da cidade referente a qual se ajusta o modelo:

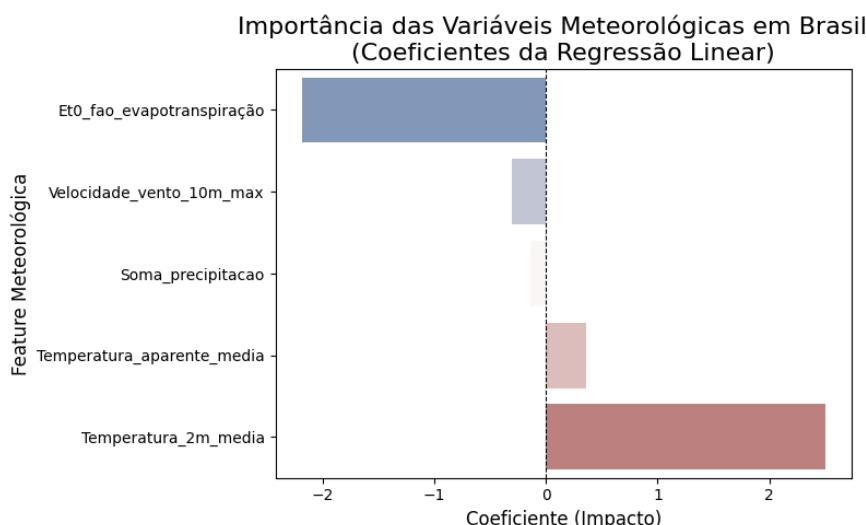


Figura 24: Importância das variáveis meteorológicas em Brasília.

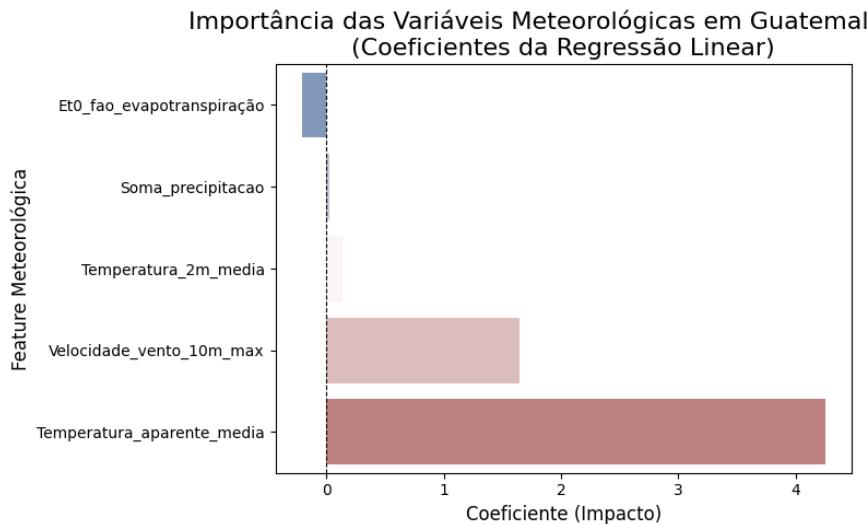


Figura 25: Importância das variáveis meteorológicas na Cidade da Guatemala.

Os demais gráficos gerados podem ser consultados no Apêndice B.6 Importância das Variáveis por Cidade.

## 4.7 Modelos Explicativos e Inferência Estatística

Nesta seção desenvolvemos um modelo explicativo para o material particulado fino (PM2.5), utilizando exclusivamente variáveis numéricas disponíveis no conjunto de dados. Foram removidos os identificadores de cidade e datas, de modo a focar na relação entre o PM2.5 e suas determinantes físico-químicas e meteorológicas.

Além da etapa descritiva, o objetivo aqui passa a ser estritamente inferencial: construir um modelo parcimonioso, estável e capaz de quantificar o efeito das covariáveis sobre o PM2.5, levando em conta a alta multicolinearidade entre as variáveis de temperatura.

### 4.7.1 Estratégia Geral de Modelagem

Para lidar adequadamente com a multicolinearidade e selecionar apenas as variáveis com contribuição efetiva, adotamos modelos lineares penalizados do tipo LASSO. A penalização L1 tem a propriedade de *zerar* coeficientes redundantes, o que produz um modelo mais simples, interpretável e robusto.

As covariáveis incluíam: poluentes atmosféricos (PM10, CO, NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub>), temperaturas reais e aparentes em diferentes medidas (mínima, máxima e média), precipitação total, velocidade máxima do vento e evapotranspiração (Et0). O conjunto completo de preditores apresentava forte redundância, especialmente entre as oito variáveis de temperatura.

#### 4.7.2 Abordagem 1: LASSO com Validação Cruzada (LassoCV)

Nesta abordagem utilizamos o LassoCV, que implementa a penalização L1 por meio do algoritmo de Descida de Coordenadas e calcula o *caminho de regularização* de forma contínua.

- **Validação Cruzada:** o LassoCV selecionou automaticamente o parâmetro de regularização  $\alpha$  que minimiza o erro de predição, utilizando CV com 5 folds.
- **Escalonamento:** todas as variáveis foram padronizadas (Z-score), passo essencial para modelos penalizados.
- **Resultado da Penalização:** o LASSO eliminou 6 variáveis, incluindo todas as formas redundantes de temperatura média e aparente.
- **Desempenho:** o modelo apresentou  $R^2 = 0,9912$  no conjunto de teste.
- **Multicolinearidade Pós-LASSO:** o Fator de Inflação da Variância (VIF) foi calculado apenas nas variáveis que permaneceram no modelo. Todos os VIFs ficaram abaixo de 10, indicando que a multicolinearidade foi efetivamente controlada.

De maneira geral, o LassoCV produziu um modelo estatisticamente estável, com excelente desempenho preditivo e forte capacidade de identificar variáveis redundantes. A interpretação dos coeficientes torna-se mais direta e confiável.

#### 4.7.3 Abordagem 2: Grid Search com Multiprocessing

Como comparação, aplicamos também uma estratégia de **força bruta** via GridSearchCV, variando 50 valores de penalização  $\alpha$  em escala logarítmica.

- **Paralelização:** a busca utilizou processamento paralelo (`n_jobs = -1`), treinando diferentes modelos simultaneamente.
- **Resultado:** o melhor  $\alpha$  encontrado foi o menor possível no grid (0.0001), o que levou o modelo a se comportar quase como uma regressão OLS.
- **Consequência:** nenhuma variável foi zerada, resultando em 14 preditores mantidos e levando a valores extremamente altos de VIF, evidenciando grave multicolinearidade.
- **Desempenho:** apesar de o  $R^2$  final também ter sido igual a 0,9912, o modelo não é adequado do ponto de vista explicativo, pois apresenta instabilidade estatística.

Assim, embora o Grid Search obtenha desempenho preditivo equivalente, ele falha em capturar a estrutura de redundância entre as variáveis, confirmando que a seleção automática e contínua do LASSO é superior para fins explicativos.

#### 4.7.4 Abordagem 3: Busca Bayesiana com Optuna

A terceira estratégia consistiu em otimizar o parâmetro de regularização  $\alpha$  por meio de **Busca Bayesiana** utilizando a biblioteca Optuna. Diferentemente do Grid Search, que avalia exaustivamente todos os pontos de um grid pré-definido, a Busca Bayesiana atualiza iterativamente a distribuição dos hiperparâmetros, concentrando-se nas regiões mais promissoras do espaço de busca.

- **Escalonamento:** assim como nas abordagens anteriores, todos os predores foram padronizados via Z-score, garantindo compatibilidade com a penalização L1.
- **Função Objetivo:** para cada *trial*, o Optuna sugeriu valores de  $\alpha$  em escala logarítmica no intervalo  $[10^{-5}, 10^2]$ . O desempenho foi avaliado via RMSE médio de uma validação cruzada com 5 folds.
- **Otimização:** foram executados 150 *trials*, nos quais o Optuna ajustou iterativamente o modelo, atualizando as distribuições probabilísticas internas para explorar regiões de melhor desempenho.
- **Resultado:** o valor ótimo encontrado foi  $\alpha^* \approx 1 \times 10^{-5}$ , o que resultou no menor RMSE médio obtido entre todas as iterações.
- **Comportamento do Modelo:** a penalização ótima encontrada é extremamente baixa, indicando que o modelo se aproxima de uma regressão OLS, porém mantendo a estabilidade proporcionada pela penalização L1.
- **Seleção de Variáveis:** embora a regularização seja fraca, o LASSO ainda eliminou algumas variáveis redundantes, reduzindo a dimensionalidade sem perda de desempenho.
- **Desempenho:** o modelo apresentou o mesmo  $R^2 = 0,9912$  observado nas abordagens anteriores, porém com um hiperparâmetro otimizado de forma estatisticamente mais eficiente.

Em síntese, a Busca Bayesiana com Optuna forneceu o processo mais eficiente para a seleção do hiperparâmetro  $\alpha$ , evitando a avaliação exaustiva do Grid Search e identificando rapidamente a região ótima do espaço de busca. O modelo resultante mantém

excelente desempenho e apresenta seletividade moderada, equilibrando explicabilidade e precisão.

## 5 Discussão

Os resultados obtidos demonstram que a qualidade do ar nas capitais latino-americanas apresenta forte variabilidade espacial e temporal, condicionada por fatores meteorológicos, padrões urbanos e características geográficas locais. Observou-se que cidades em regiões andinas e em áreas com ventilação atmosférica limitada apresentaram maiores concentrações de PM<sub>2.5</sub> e NO<sub>2</sub>, enquanto capitais litorâneas, mais ventiladas e úmidas, exibiram níveis mais moderados de poluentes.

A comparação entre Brasília e as demais capitais indicou diferenças estatisticamente significativas nas médias de poluição para diversos poluentes, sugerindo que o padrão atmosférico relativamente seco e a baixa ventilação sazonal da capital brasileira contribuem para concentrações mais elevadas em determinados períodos. O teste F evidenciou heterogeneidade de variâncias, justificando o uso do teste de Welch, que confirmou discrepâncias relevantes nos níveis médios, especialmente para PM<sub>2.5</sub>.

As análises de correlação reforçaram o papel determinante das variáveis meteorológicas na dinâmica dos poluentes. Temperatura, umidade e velocidade do vento apresentaram relações consistentes com concentrações de PM<sub>2.5</sub> e O<sub>3</sub>, refletindo mecanismos conhecidos de dispersão e formação fotoquímica. A forte dependência temporal capturada pelo lag de PM<sub>2.5</sub> indica persistência diária das condições de poluição.

O modelo de regressão múltipla apresentou desempenho satisfatório, com  $R^2$  elevado e RMSE compatível com a variabilidade observada. A inclusão de variáveis categóricas (cidade, mês e dia da semana) aumentou a capacidade explicativa do modelo, evidenciando a influência tanto de padrões urbanos quanto de fatores sazonais. Já a convergência dos métodos de seleção de hiperparâmetros (LassoCV, Grid Search e Optuna) para o modelo preditivo indica robustez no processo de modelagem.

## 6 Conclusão

A análise realizada permitiu caracterizar de forma abrangente a qualidade do ar nas capitais latino-americanas, destacando diferenças estruturais relevantes para a compreensão dos desafios ambientais enfrentados pela região. Verificou-se que Brasília possui padrões de poluição diferenciados em relação às demais capitais, com diferenças estatisticamente significativas para diversos poluentes.

Variáveis meteorológicas, especialmente umidade e vento, mostraram-se fundamentais para explicar a variabilidade das concentrações de PM<sub>2.5</sub>, enquanto características

temporais e específicas de cada cidade complementaram a capacidade preditiva do modelo. O modelo proposto demonstrou boa performance e pode servir como base para aprimoramentos futuros em termos de previsão operacional e formulação de políticas públicas.

De modo geral, os resultados reforçam a importância de integrar dados meteorológicos, históricos e geográficos em estudos de qualidade do ar, particularmente em um contexto continental diverso como o latino-americano. Para a COP30, tais análises contribuem para posicionar o Brasil no debate regional sobre poluição atmosférica e estratégias de mitigação.

## 7 Referências

### Referências

- [1] World Health Organization. *Global Air Quality Guidelines*. WHO, 2021
- [2] Kaggle. *Latin America Weather and Air Quality Data*. Disponível em: <https://www.kaggle.com>. Acesso em: XX.

## Apêndice A

### Testes de Igualdade de Variâncias

Este apêndice apresenta os resultados completos dos testes F aplicados para comparar as variâncias dos níveis de poluentes atmosféricos entre Brasília e o conjunto das demais capitais. Os valores não foram incluídos no corpo do relatório por questões de concisão, mas são disponibilizados aqui para fins de transparência metodológica.

#### A.1 PM10

---

Variância de Brasília ( $s_1^2$ )	72,98
Variância das demais capitais ( $s_2^2$ )	295,17
Estatística F observada	0.25
Região de aceitação	[0.89, 1.12]
Conclusão	Rejeita-se $H_0$ (variâncias diferentes)

---

**A.2 PM<sub>2.5</sub>**


---

Variância de Brasília ( $s_1^2$ )	35,50
Variância das demais capitais ( $s_2^2$ )	141,11
Estatística F observada	0.25
Região de aceitação	[0.89, 1.12]
Conclusão	Rejeita-se $H_0$ (variâncias diferentes)

---

**A.3 Dióxido de Nitrogênio (NO<sub>2</sub>)**


---

Variância de Brasília ( $s_1^2$ )	317,33
Variância das demais capitais ( $s_2^2$ )	628,18
Estatística F observada	0.51
Região de aceitação	[0.89, 1.12]
Conclusão	Rejeita-se $H_0$ (variâncias diferentes)

---

**A.4 Dióxido de Enxofre (SO<sub>2</sub>)**


---

Variância de Brasília ( $s_1^2$ )	1,75
Variância das demais capitais ( $s_2^2$ )	194,68
Estatística F observada	0.01
Região de aceitação	[0.89, 1.12]
Conclusão	Rejeita-se $H_0$ (variâncias diferentes)

---

**A.5 Ozônio (O<sub>3</sub>)**


---

Variância de Brasília ( $s_1^2$ )	448,63
Variância das demais capitais ( $s_2^2$ )	1,071,74
Estatística F observada	0.42
Região de aceitação	[0.89, 1.12]
Conclusão	Rejeita-se $H_0$ (variâncias diferentes)

---

## Apêndice B

### Gráficos Complementares

Este apêndice reúne as visualizações que complementam as análises estatísticas apresentadas no relatório principal. Os gráficos são organizados por tipo e por poluente, facilitando a consulta e a comparação entre padrões observados.

## B.1 Gráficos de Violino

$\text{SO}_2$

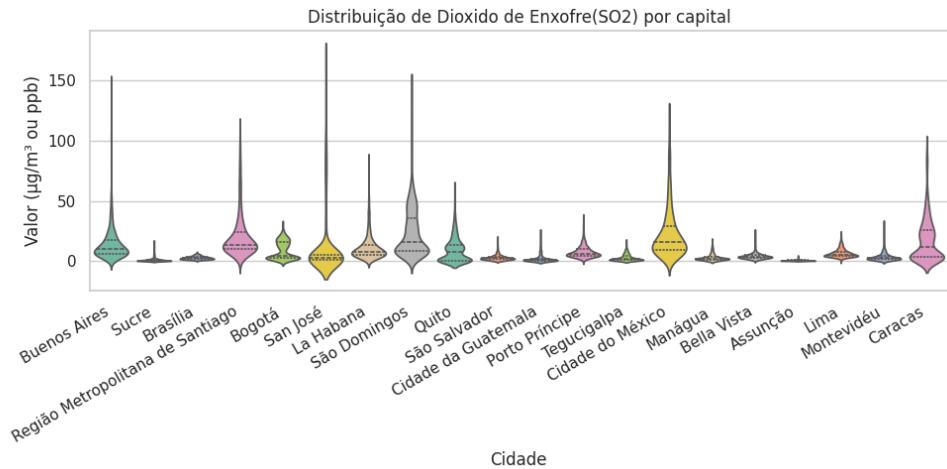


Figura 26: Distribuição dos valores de  $\text{SO}_2$  (gráfico de violino).

## B.2 Gráficos de Limites

$\text{PM10}$

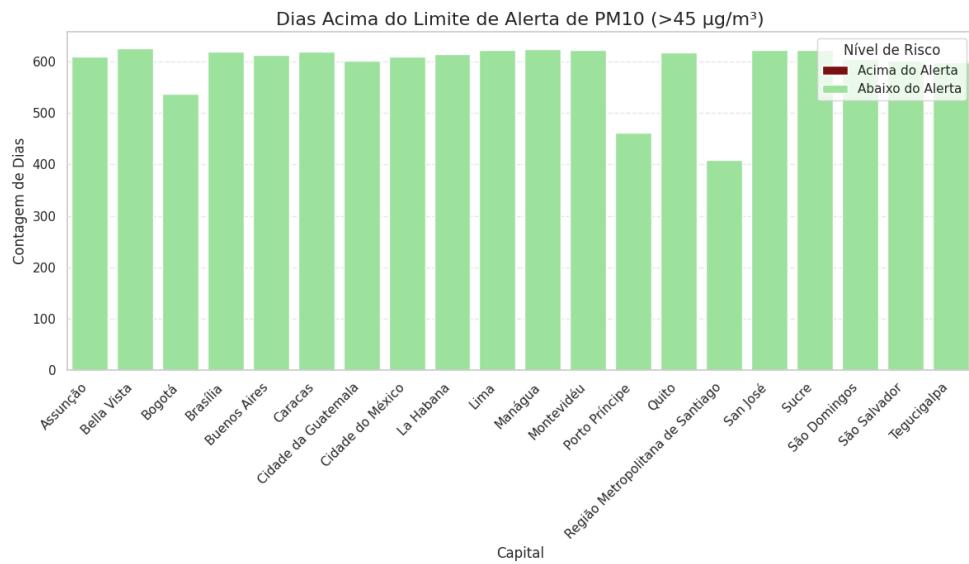
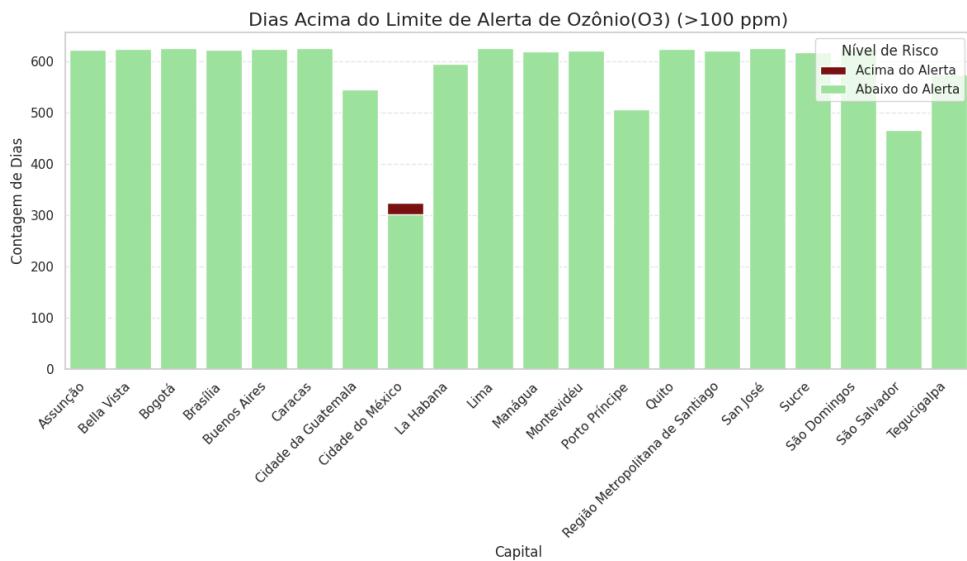
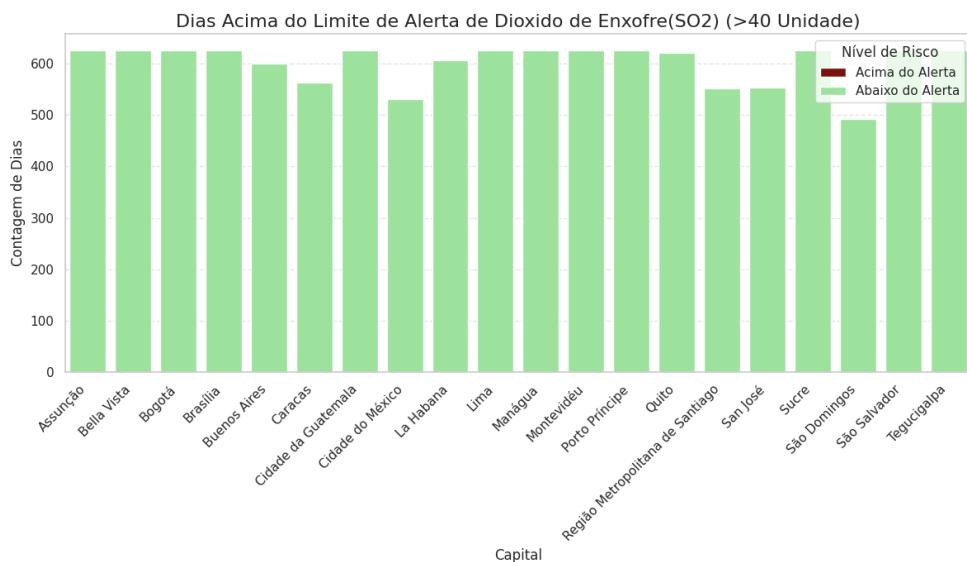


Figura 27: Gráfico de limites para  $\text{PM10}$ .

**O<sub>3</sub>**Figura 28: Gráfico de limites para O<sub>3</sub>.**SO<sub>2</sub>**Figura 29: Gráfico de limites para SO<sub>2</sub>.

### B.3 Boxplots

CO

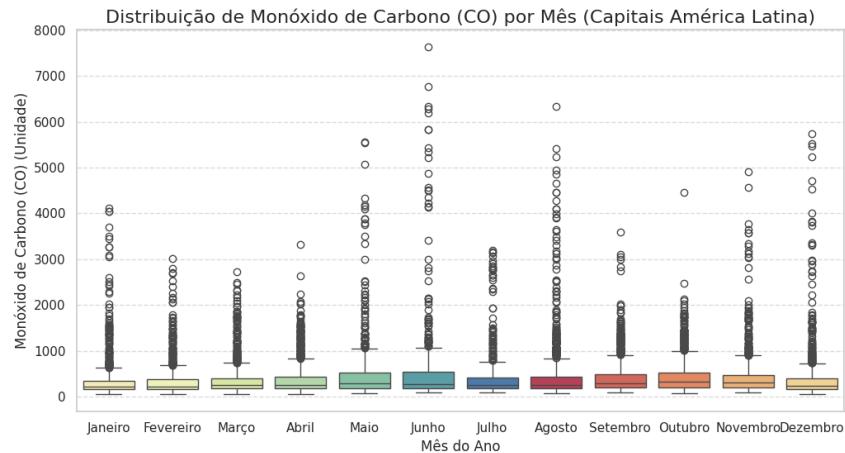


Figura 30: Boxplot dos valores de CO.

NO<sub>2</sub>

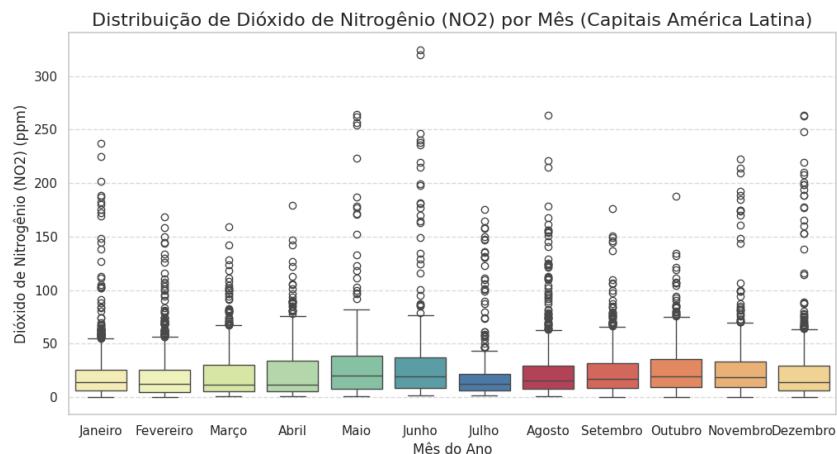


Figura 31: Boxplot dos valores de NO<sub>2</sub>.

**SO<sub>2</sub>**

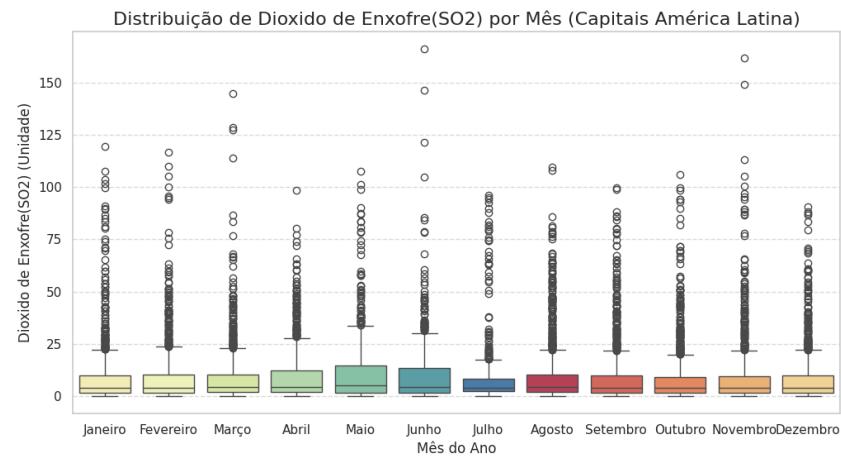


Figura 32: Boxplot dos valores de SO<sub>2</sub>.

## B.4 Calendários de Calor

Ano de 2022

**SO2**

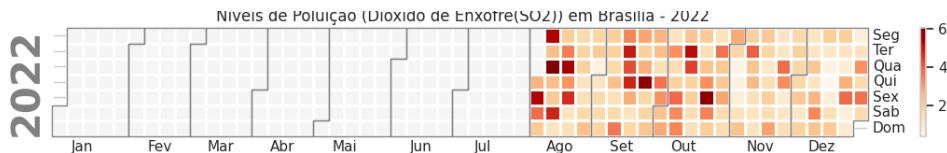


Figura 33: Calendário de calor da concentração de Dióxido de Enxofre (SO<sub>2</sub>) em Brasília ao longo de 2022.

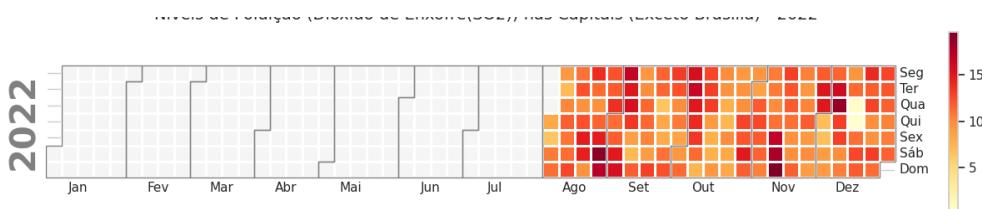


Figura 34: Calendário de calor da concentração de Dióxido de Enxofre (SO<sub>2</sub>) nas demais capitais ao longo de 2022.

**NO2**

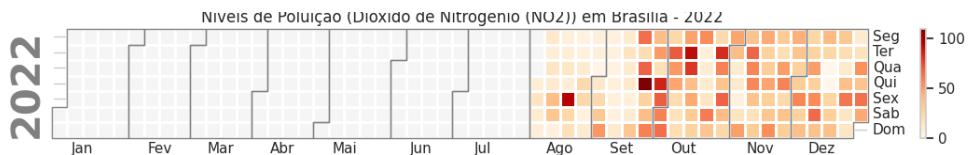


Figura 35: Calendário de calor da concentração de Dióxido de Nitrogênio (NO<sub>2</sub>) em Brasília ao longo de 2022.

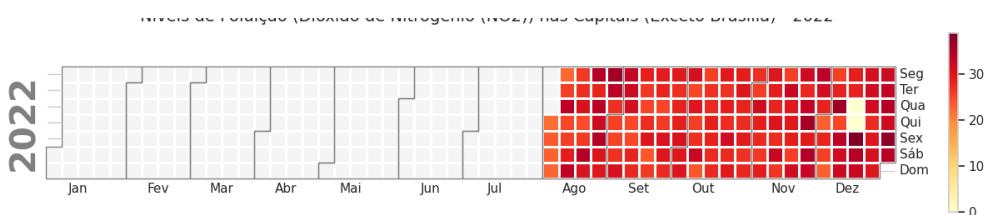


Figura 36: Calendário de calor da concentração de Dióxido de Nitrogênio (NO<sub>2</sub>) nas demais capitais ao longo de 2022.

## Ano de 2024

### SO2

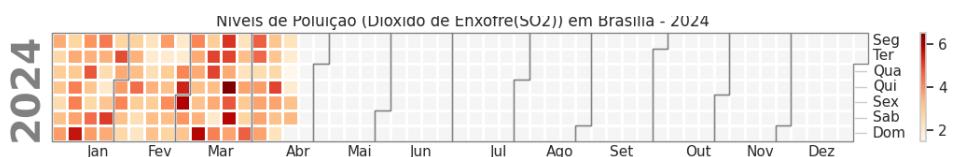


Figura 37: Calendário de calor da concentração de Dióxido de Enxofre (SO<sub>2</sub>) em Brasília ao longo de 2024.

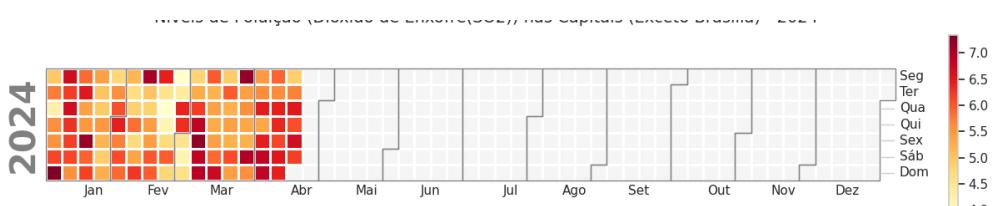


Figura 38: Calendário de calor da concentração de Dióxido de Enxofre (SO<sub>2</sub>) nas demais capitais ao longo de 2024.

### NO2

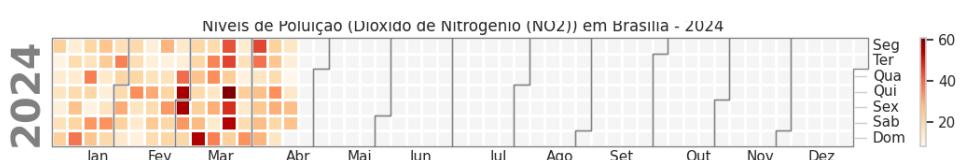


Figura 39: Calendário de calor da concentração de Dióxido de Nitrogênio (NO<sub>2</sub>) em Brasília ao longo de 2024.

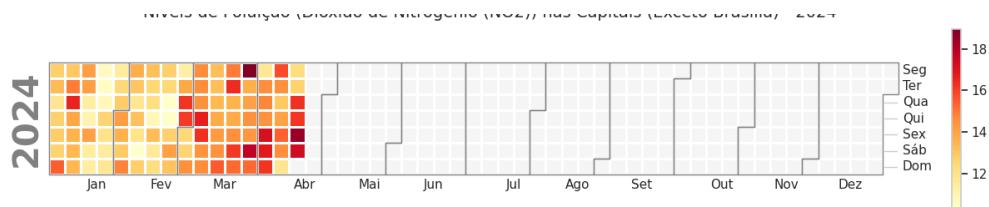


Figura 40: Calendário de calor da concentração de Dióxido de Nitrogênio (NO<sub>2</sub>) nas demais capitais ao longo de 2024.

**CO**

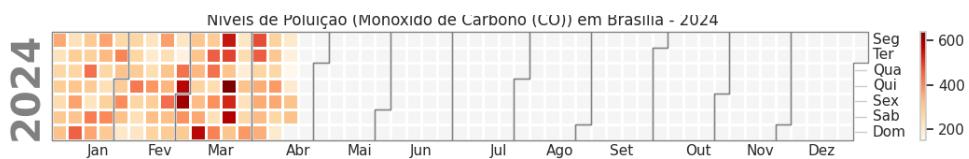


Figura 41: Calendário de calor da concentração de Monóxido de Carbono (CO) em Brasília ao longo de 2024.

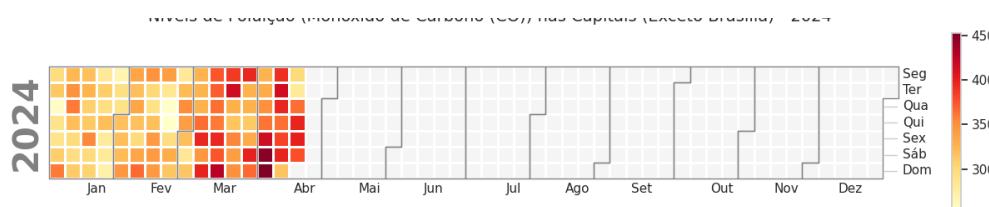


Figura 42: Calendário de calor da concentração de Monóxido de Carbono (CO) nas demais capitais ao longo de 2024.

## B.5 Relação Entre Variáveis Climáticas e Meteorológicas

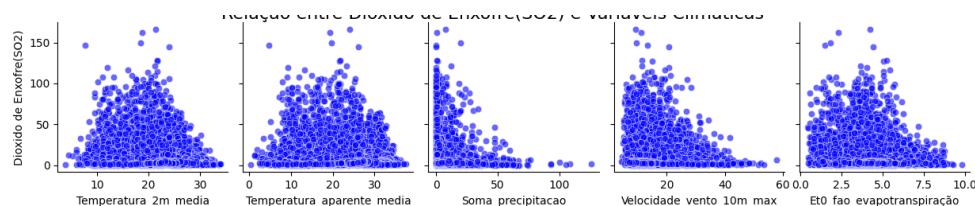


Figura 43: Relação entre o Dióxido de Enxofre (SO<sub>2</sub>) e variáveis meteorológicas.

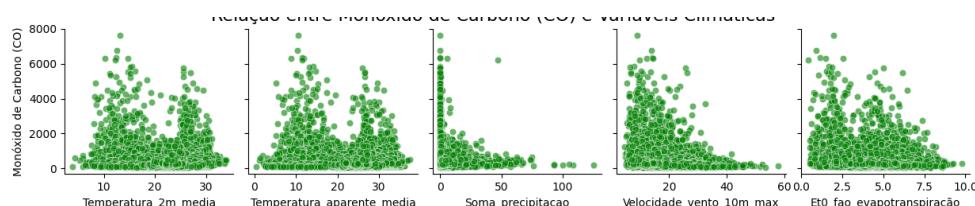


Figura 44: Relação entre Monóxido de Carbono (CO) e variáveis meteorológicas.

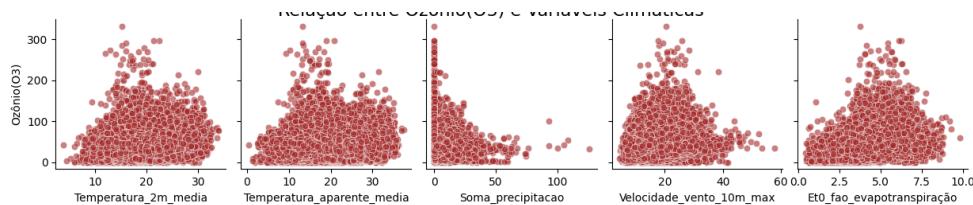


Figura 45: Relação entre Ozônio (O3) e variáveis meteorológicas.

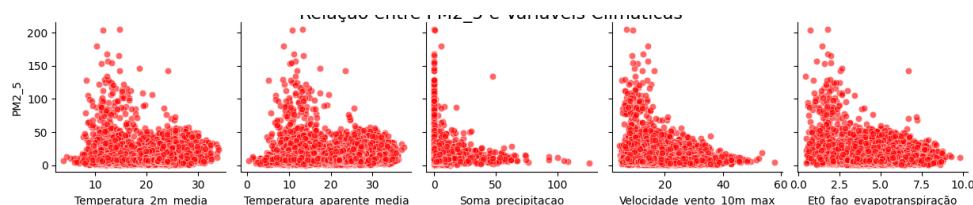


Figura 46: Relação entre PM<sub>2.5</sub> e variáveis meteorológicas.

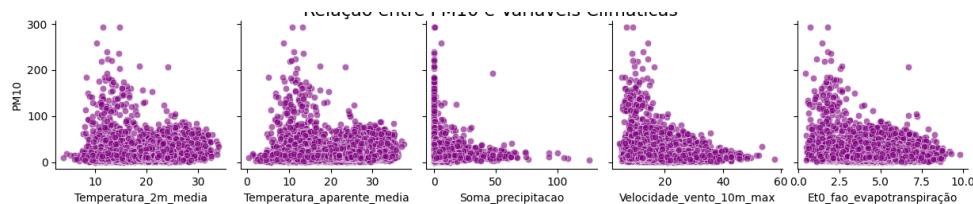


Figura 47: Relação entre PM<sub>10</sub> e variáveis meteorológicas.

## B.6 Importância das Variáveis por Cidade

Os gráficos a seguir apresentam a importância das variáveis meteorológicas estimadas por Regressão Linear Múltipla para cada cidade analisada, exceto Brasília e Cidade da Guatemala, cujos resultados já foram discutidos no corpo principal do relatório.

## Asuncion

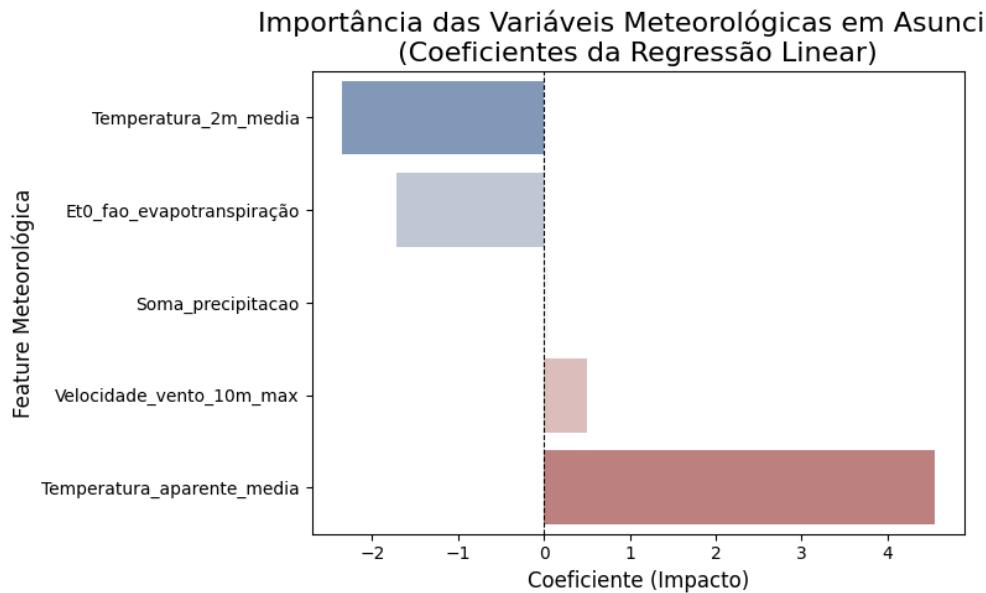


Figura 48: Importância das variáveis meteorológicas – Asuncion (Paraguai).

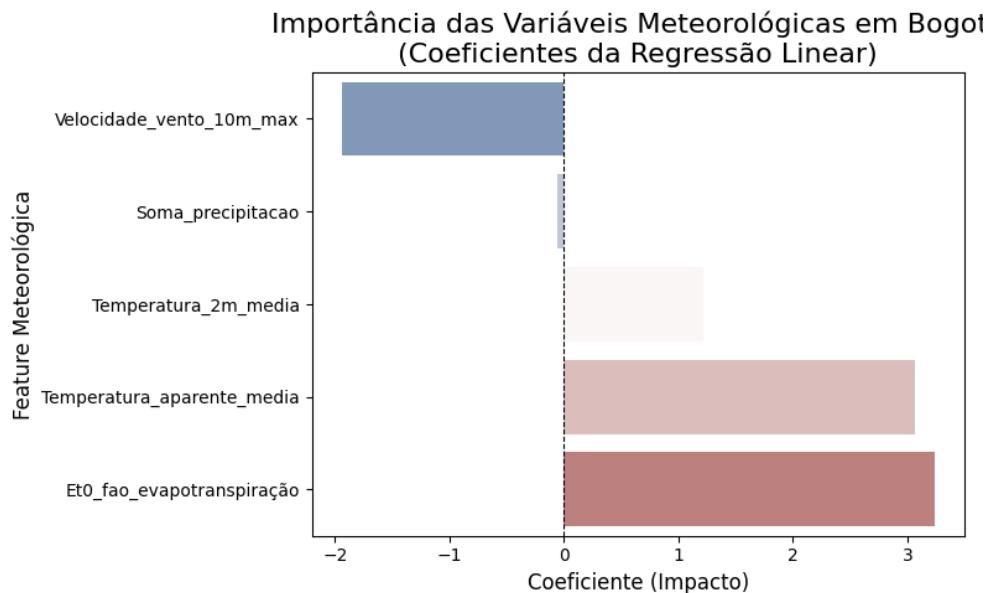


Figura 49: Importância das variáveis meteorológicas – Bogotá (Colômbia).

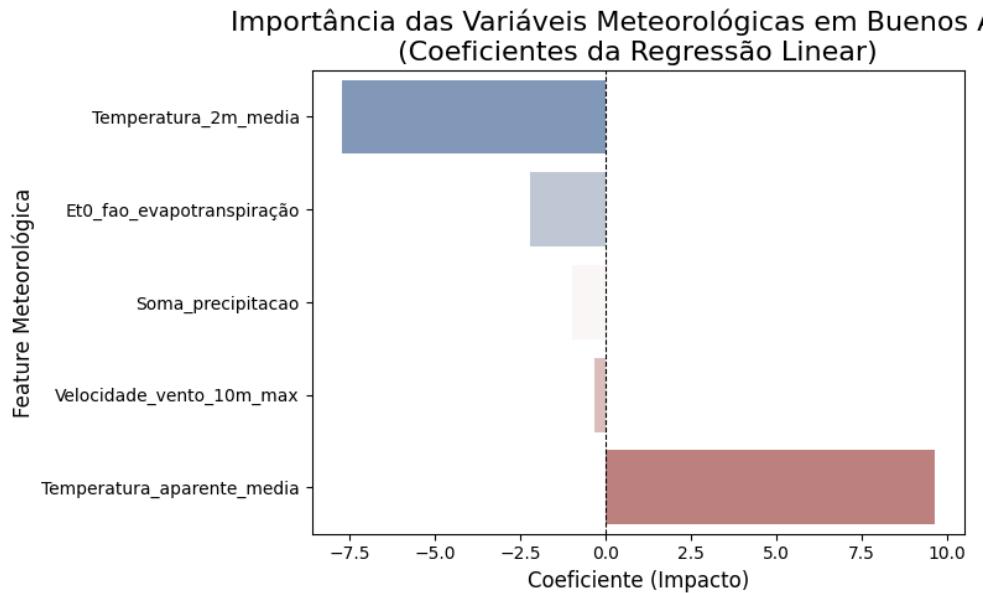


Figura 50: Importância das variáveis meteorológicas – Buenos Aires (Argentina).

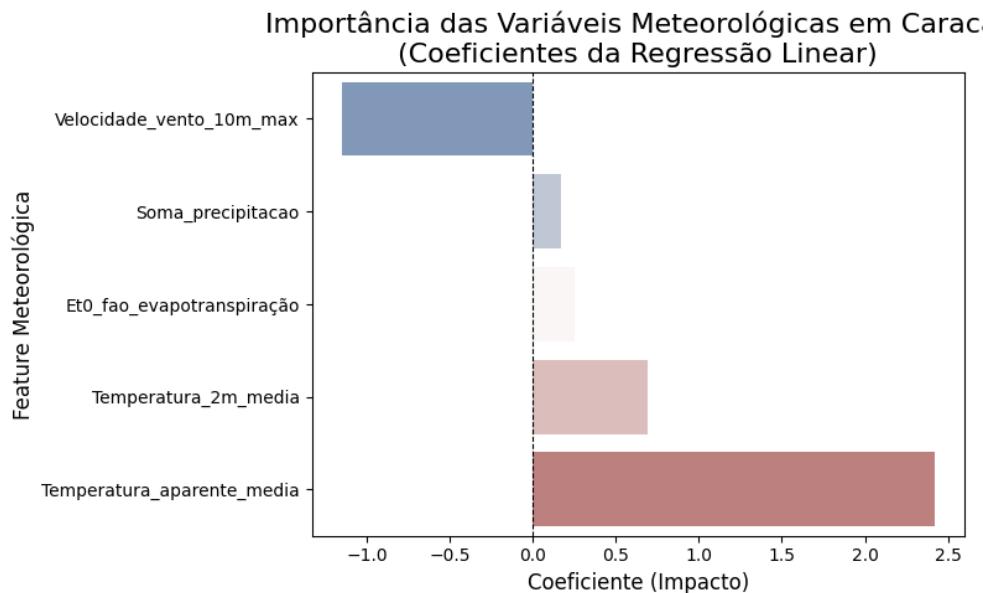


Figura 51: Importância das variáveis meteorológicas – Caracas (Venezuela).

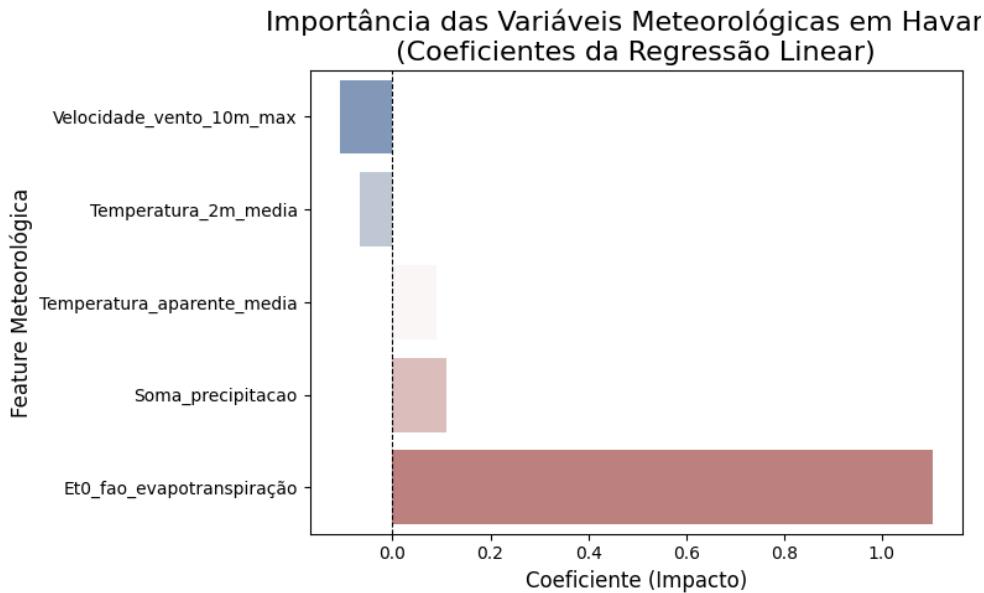


Figura 52: Importância das variáveis meteorológicas – Havana (Cuba).

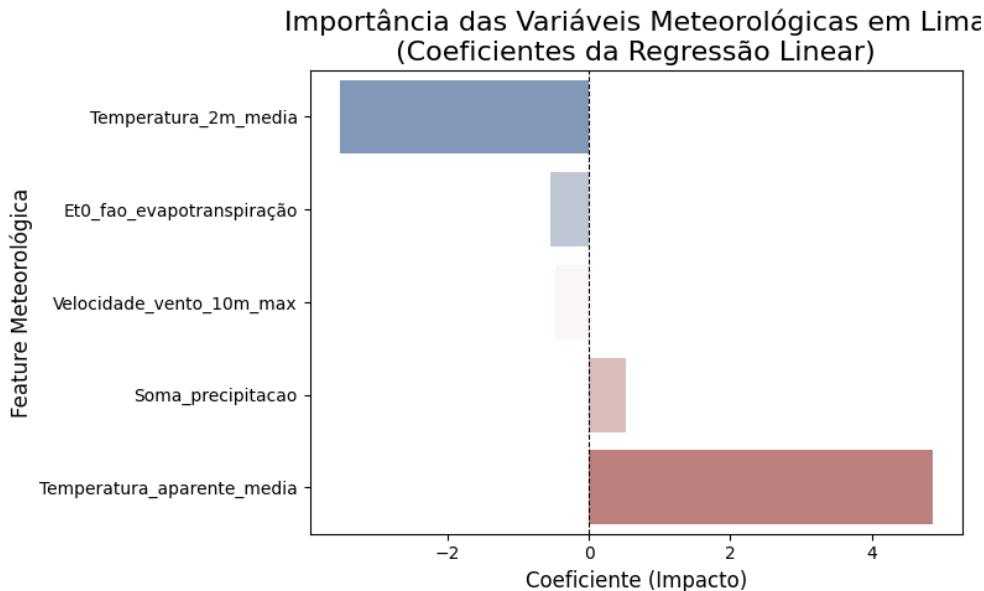


Figura 53: Importância das variáveis meteorológicas – Lima (Peru).

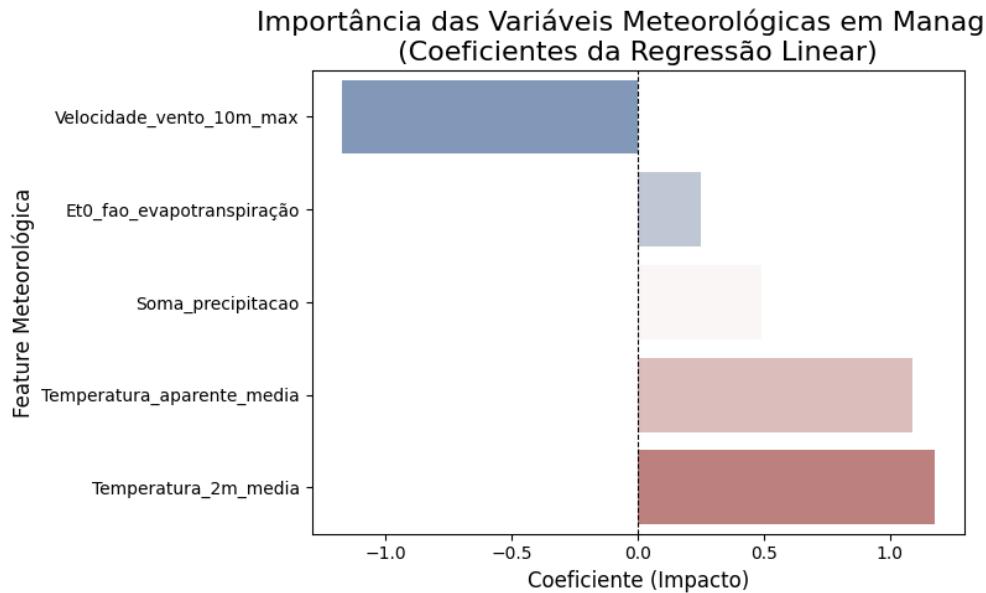


Figura 54: Importância das variáveis meteorológicas – Managuá (Nicarágua).

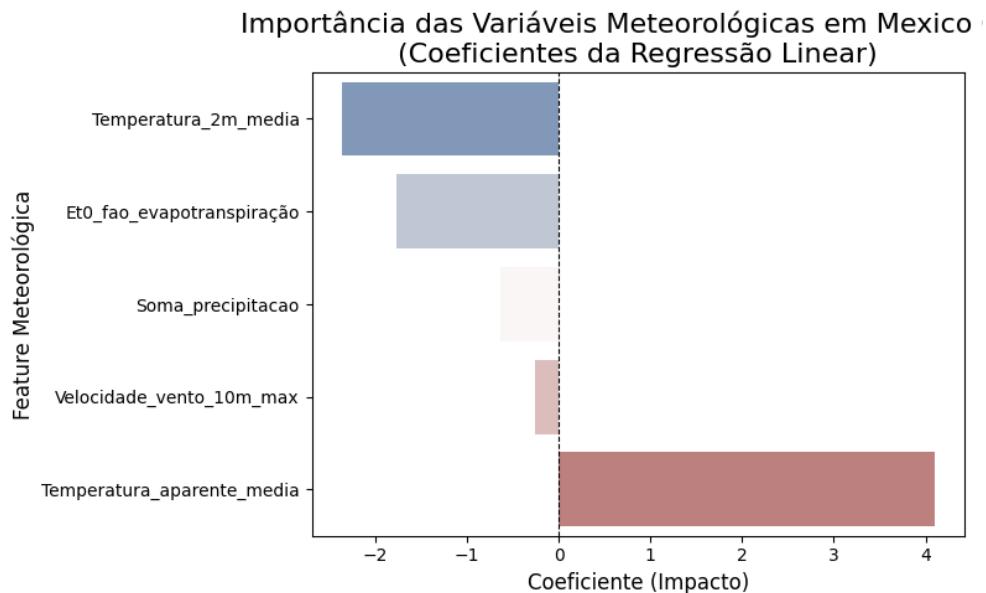


Figura 55: Importância das variáveis meteorológicas – Cidade do México (México).

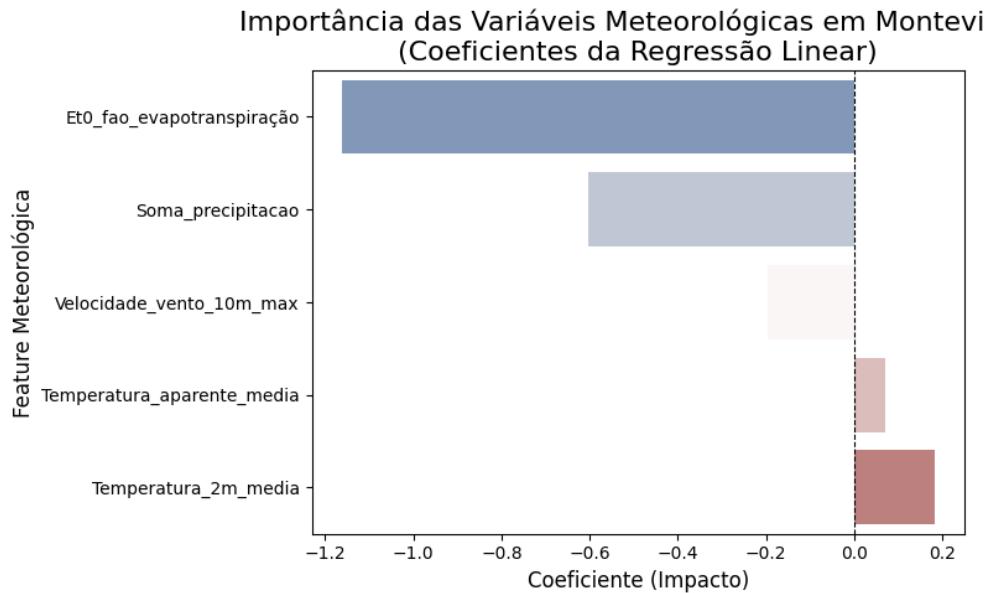


Figura 56: Importância das variáveis meteorológicas – Montevideo (Uruguai).

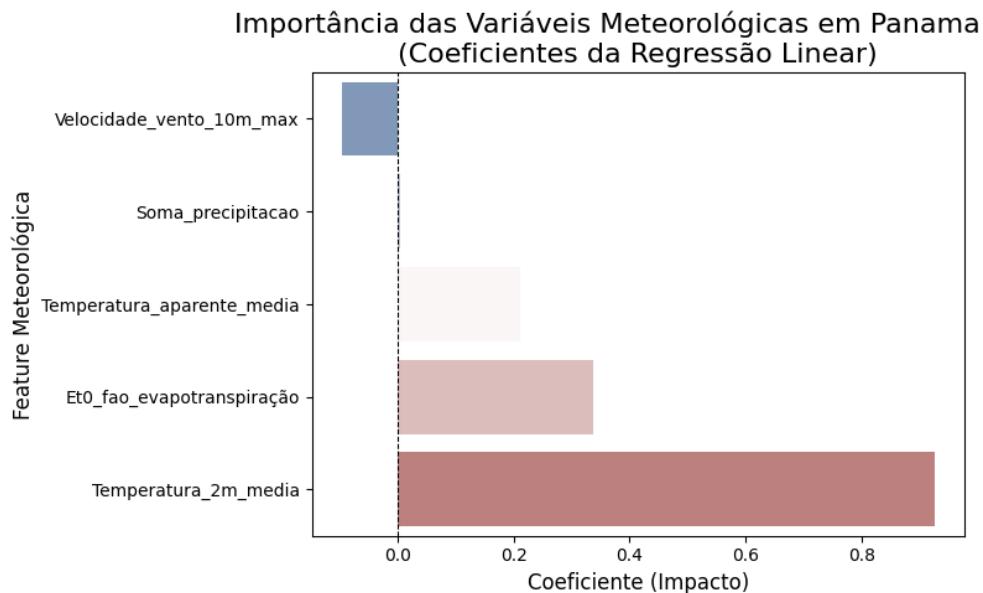


Figura 57: Importância das variáveis meteorológicas – Cidade do Panamá (Panamá).

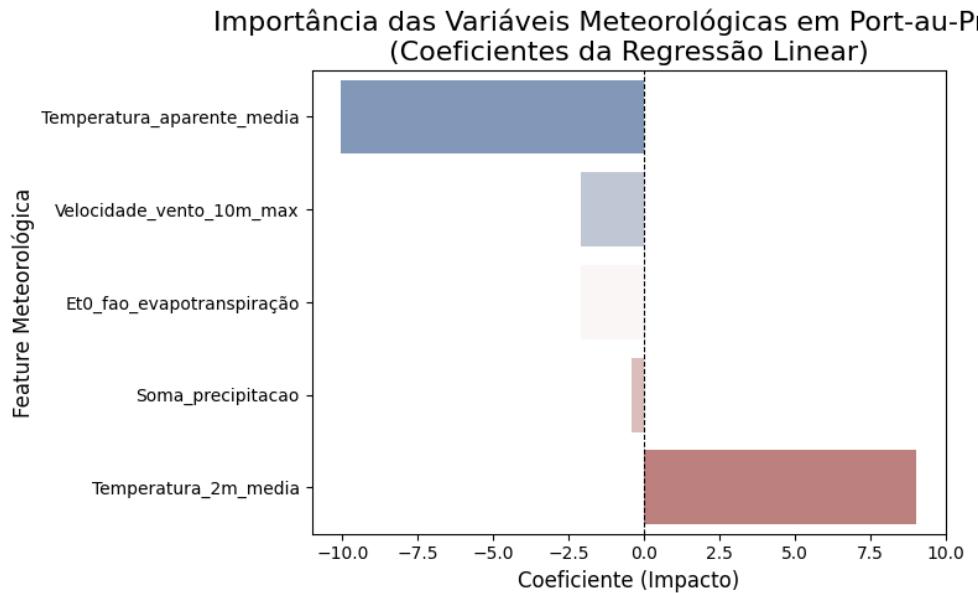


Figura 58: Importância das variáveis meteorológicas – Porto Príncipe (Haiti).

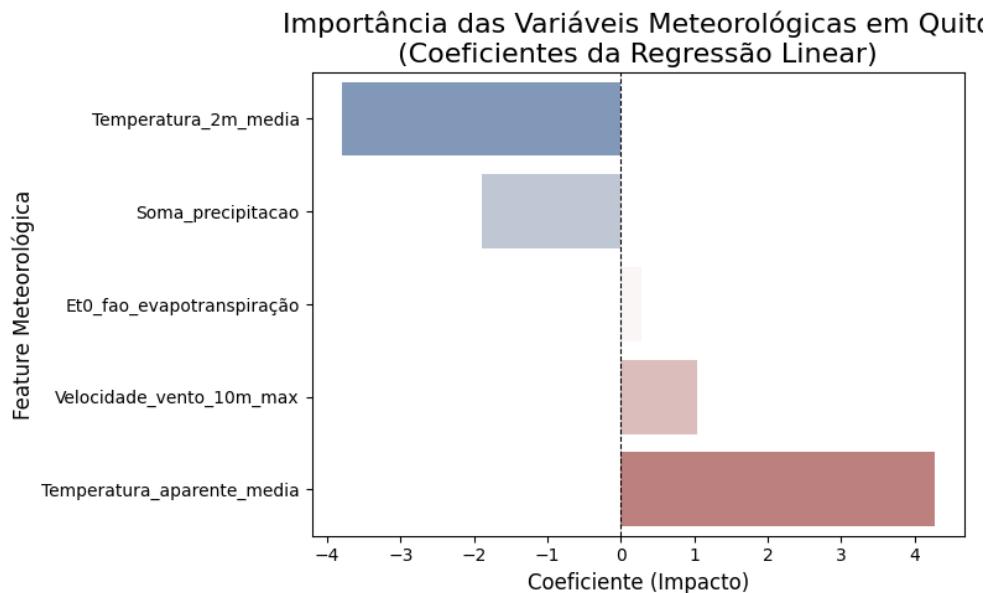


Figura 59: Importância das variáveis meteorológicas – Quito (Equador).

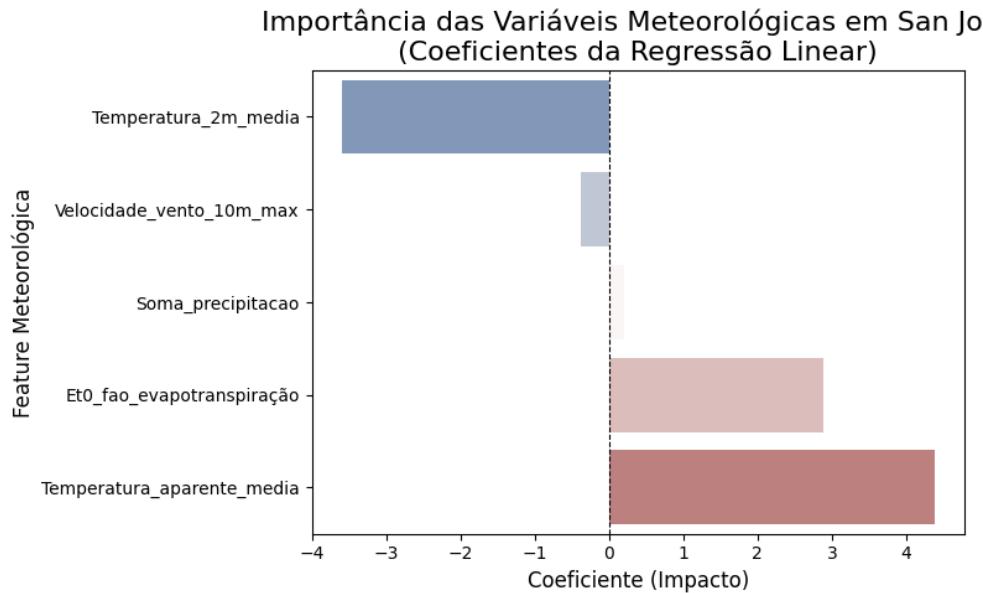


Figura 60: Importância das variáveis meteorológicas – San Jose (Costa Rica).

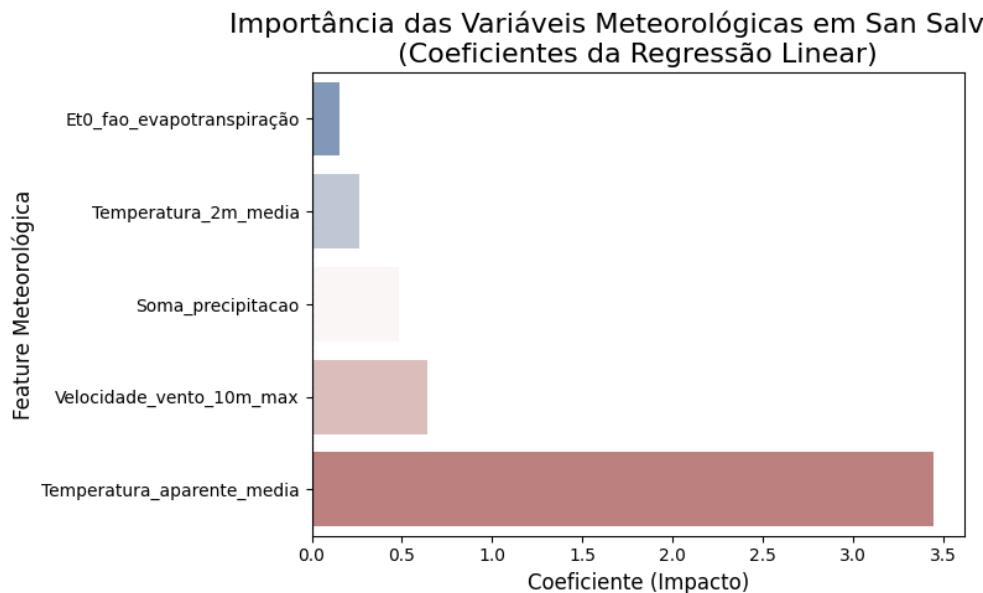


Figura 61: Importância das variáveis meteorológicas – San Salvador (El Salvador).

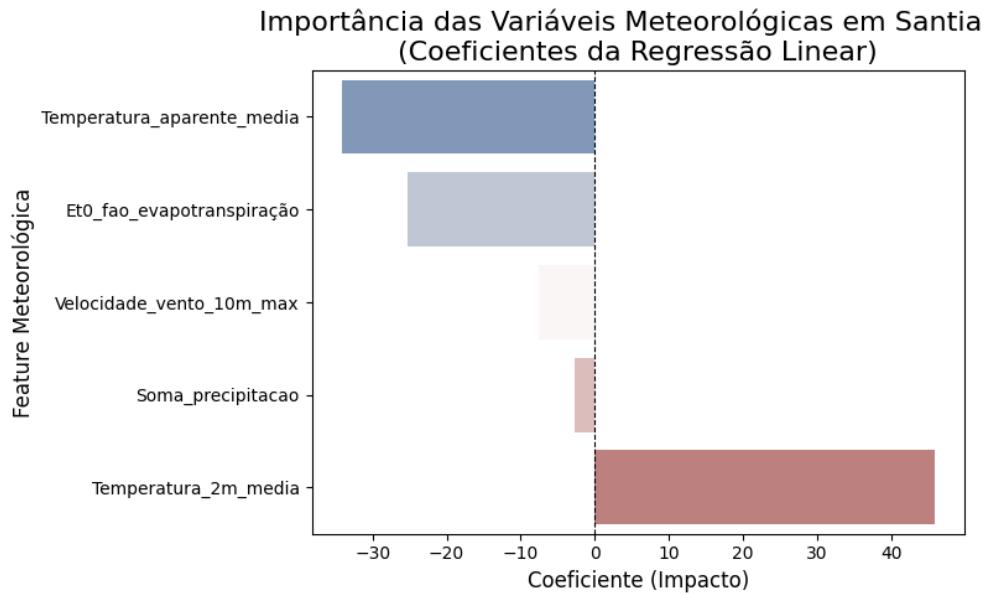


Figura 62: Importância das variáveis meteorológicas – Santiago (Chile).

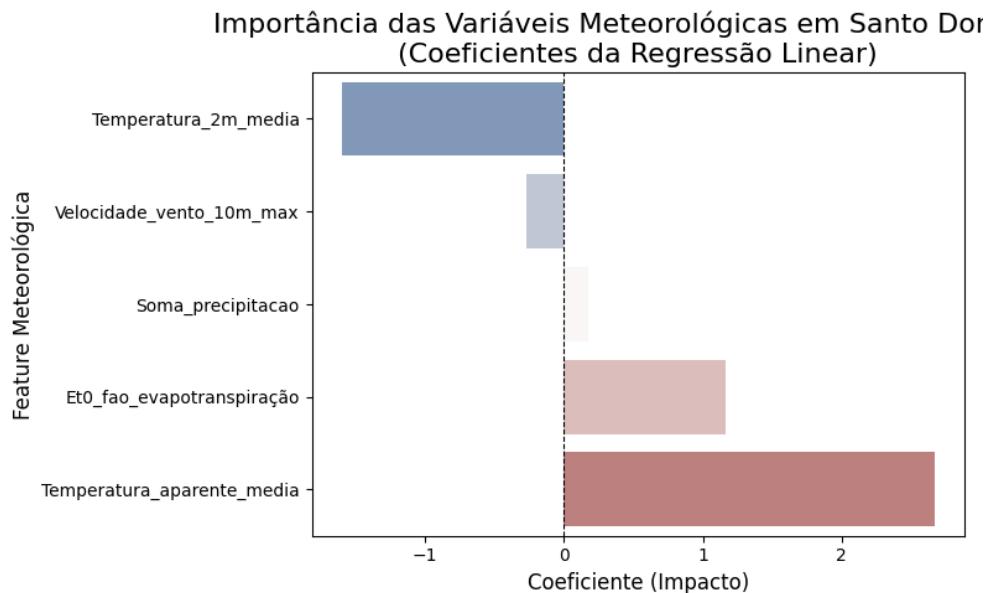


Figura 63: Importância das variáveis meteorológicas – Santo Domingo (República Dominicana).

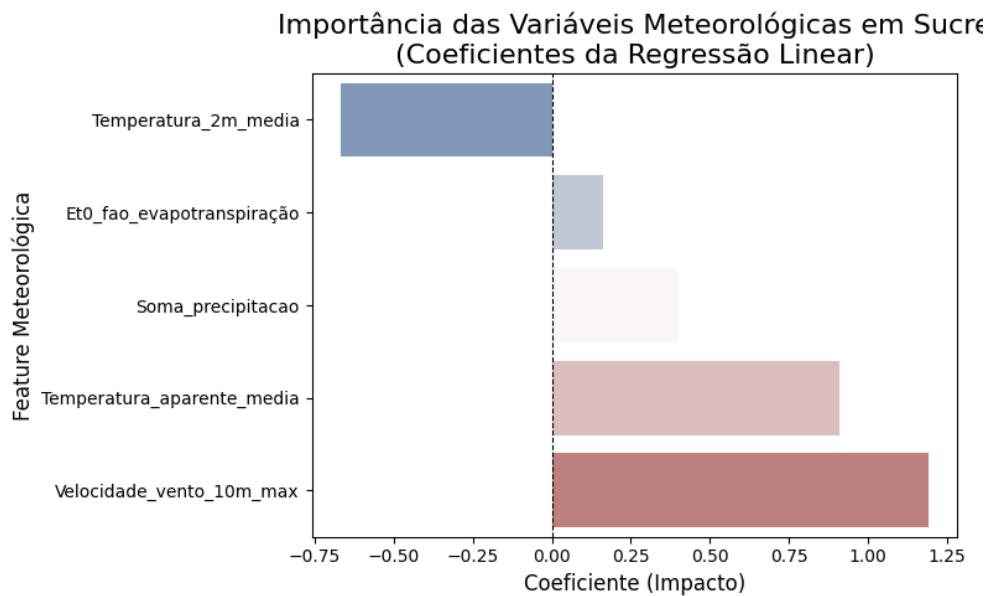


Figura 64: Importância das variáveis meteorológicas – Sucre (Bolívia).

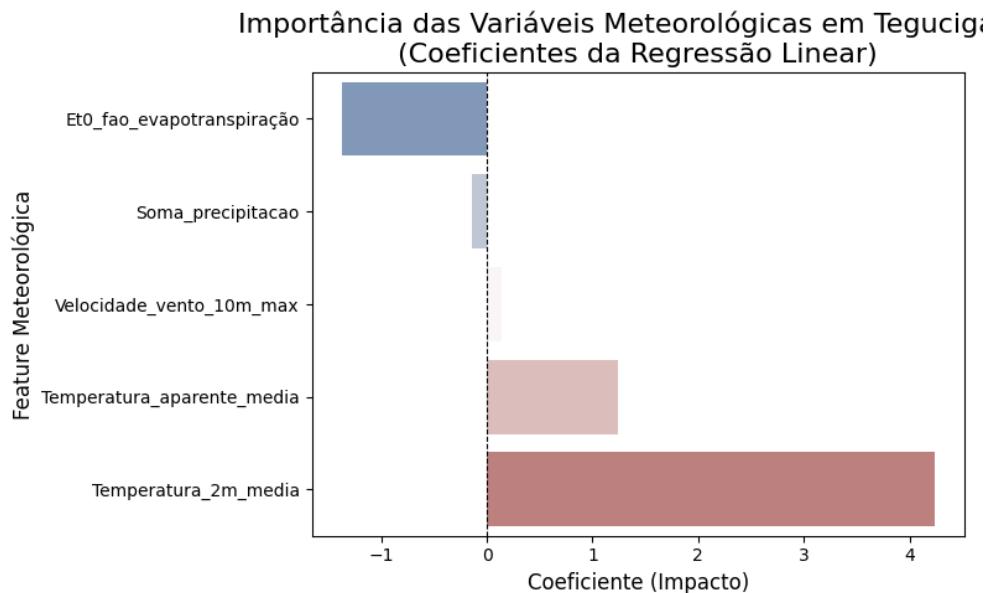


Figura 65: Importância das variáveis meteorológicas – Tegucigalpa (Honduras).