

Análise de Dados Públicos sobre Qualidade do Ar em Grande Cidades no Brasil

Jhonata Silveira Dias¹, Lucca Oliveira Vasconcelos de Faria¹,
Pedro Henrique Moreira Caixeta Ferreira¹

¹Instituto de Ciências Exatas e Informática
Departamento de Engenharia de Software
Pontifícia Universidade Católica de Minas Gerais (PUC-MG)
CEP 30535-000 – Belo Horizonte – MG – Brazil

{jhonata.dias, lovfarria}@sga.pucminas.br, phmcf13@gmail.com

Abstract. *This study assesses air quality in major Brazilian cities using open data from the OpenAQ platform. Significant disparities in sensor coverage and irregular data collection periods limited comparisons across municipalities. Nevertheless, seasonal patterns and exceedances of WHO guideline limits were identified, indicating persistent air pollution issues in monitored cities. The findings highlight the need to expand monitoring efforts and strengthen air-quality-related public policies.*

Resumo. *Este estudo avalia a qualidade do ar em grandes cidades brasileiras utilizando dados abertos da plataforma OpenAQ. A análise mostrou forte desigualdade na cobertura de sensores e irregularidade nos períodos de coleta, restringindo comparações entre municípios. Ainda assim, observaram-se padrões sazonais e episódios de excedência aos limites recomendados pela OMS, indicando que problemas relevantes de poluição atmosférica persistem nas cidades monitoradas. Os resultados reforçam a necessidade de ampliar o monitoramento e fortalecer políticas públicas de qualidade do ar.*

1. Introdução

A poluição atmosférica é um dos maiores desafios ambientais contemporâneos. Ela impacta diretamente a saúde da população, a sustentabilidade das cidades e a estabilidade climática do planeta. Entre os poluentes mais estudados estão o material particulado fino (PM_{2.5}), o material particulado inalável (PM₁₀), SO₂, NO₂, CO e O₃.

Nas últimas décadas, a urbanização acelerada, o aumento do tráfego veicular e as queimadas, especialmente nas regiões Centro-Oeste e Norte do Brasil, agravaram a exposição da população a esses poluentes [Jornal 2024]. A Organização Mundial da Saúde (OMS) estima que milhões de mortes anuais estejam relacionadas à má qualidade do ar, sendo uma das principais causas de doenças respiratórias e cardiovasculares [World Health Organization 2021, Silva and Santos 2021]. Diante desse cenário, torna-se urgente investigar, de forma sistemática e aberta, a situação da qualidade do ar nas cidades brasileiras, utilizando dados públicos e ferramentas modernas de análise.

O presente trabalho tem como objetivo desenvolver uma base analítica e visual (*dashboard BI*) para avaliar a qualidade do ar em cidades de grande porte brasileiras.

Como cidade de grande porte leva-se em consideração a definição do Instituto Brasileiro de Geografia e Estatística (IBGE) para cidades grandes, ou seja, cidades com mais que 500 mil habitantes [IBGE — Instituto Brasileiro de Geografia e Estatística 2008]. Os dados da qualidade do ar são providenciados pela plataforma OpenAQ, uma iniciativa global de dados abertos que integra medições de centenas de estações monitoras espalhadas pelo mundo.

2. Metodologia e Descrição da Base de Dados

2.1. Objetivo

O objetivo deste estudo é investigar os padrões espaciais e temporais da qualidade do ar em grandes cidades brasileiras, identificando diferenças regionais, tendências sazonais e a frequência com que os níveis de poluentes ultrapassam os limites recomendados pela Organização Mundial da Saúde (OMS). A análise visa oferecer um panorama comparativo entre municípios, destacando disparidades na infraestrutura de monitoramento e na regularidade da coleta.

2.2. Perguntas de Pesquisa

- RQ1 - Variação Espacial: Quais cidades apresentam os maiores valores médios de concentração de poluentes?
- RQ2 - Padrões Temporais e Sazonais: Há variações significativas nas concentrações dos poluentes ao longo dos anos ou meses?
- RQ3 - Excedência dos Limites da OMS: Qual é o percentual do ano em meses em que os níveis de poluentes ultrapassam os limites seguros definidos pela OMS?

2.3. Metrics

Para responder às perguntas de pesquisa, foram adotadas como métricas principais:

- Concentração média anual por cidade e por poluente - em unidades padronizadas ($\mu\text{g}/\text{m}^3$ e ppm para CO).
- Médias mensais dos poluentes - utilizadas para identificar padrões sazonais.
- Percentual de meses acima dos limites OMS - calculado para cada combinação cidade–poluente.
- Cobertura temporal da coleta (*percentCoverage*) - utilizada para avaliar a confiabilidade da série temporal de cada sensor.
- Número de sensores por cidade - empregado como indicador indireto da densidade da rede de monitoramento.

2.4. Fonte de Dados

Os dados utilizados neste estudo foram obtidos a partir da OpenAQ, uma plataforma global de acesso aberto que reúne medições de qualidade do ar provenientes de agências governamentais, monitoramentos acadêmicos e outras instituições certificadas. A OpenAQ disponibiliza dados padronizados de poluentes atmosféricos, incluindo PM2.5, PM10, SO₂, NO₂, CO e O₃ e outros parâmetros regulamentados.

A coleta foi realizada exclusivamente através da API oficial da OpenAQ (versão 3), que permite o acesso direto às localizações, sensores e séries temporais agregadas em diferentes escalas (horária, diária, mensal e anual). Em particular, utilizou-se os *end-points*:

- `/v3/locations` - para identificar sensores ativos em cada cidade.
- `/v3/sensors{sensor_id}/days/monthly` - para obter agregações mensais dos dados coletados.
- `/v3/sensors{sensor_id}/days/yearly` - para obter agregações anuais dos dados coletados.

2.5. Seleção das Cidades e Critério Geoespacial

Foram selecionadas todas as cidades brasileiras com mais de 500 mil habitantes, totalizando 35 municípios. Para cada cidade, utilizou-se sua coordenada central (latitude e longitude) como ponto de referência.

A API OpenAQ não fornece consultas diretamente por nome de cidade, portanto foi utilizada a busca geoespacial baseada em raio (buffer). O valor do raio utilizado foi de 12 km em torno do ponto central de cada cidade.

Esse procedimento permitiu identificar todas as estações de monitoramento próximas à região urbana principal, garantindo que apenas sensores relevantes fossem incluídos. Apesar de que em grandes regiões metropolitanas, como São Paulo, possa haver uma sobreposição das áreas, essa foi a melhor solução encontrada para coletar o máximo de dados disponíveis.

2.6. Coleta Automatizada dos Dados

A coleta foi realizada por meio de um *script* desenvolvido em Python, utilizando as bibliotecas:

- `requests` - consumo da *API*.
- `pandas` - organização tabular e limpeza dos dados.
- `dotenv` - gerenciamento seguro da chave da *API*.
- `time` - controle de chamadas para evitar *rate limit*.

O processo foi estruturado em quatro etapas principais:

- Identificação dos sensores por cidade, a partir da busca geoespacial com raio de 12 km.
- Coleta dos dados anuais e mensais de cada sensor identificado.
- Agregação por cidade, integrando dados provenientes de sensores distintos e calculando estatísticas consolidadas.
- Visualização de dados, exportação de dados e criação de gráficos pelo Power BI.

3. Resultados

Table 1. Tabela De Cobertura de Sensores

Cidade	N. Sensores	Data Início	Data Final	Cobertura Média	Anos de Cobertura Equivalente
Belém	5	2025-01-01	2026-01-01	32.0	0.32
Campinas	9	2017-01-01	2024-01-01	47.8	3.35
Duque de Caxias	20	2023-01-01	2026-01-01	55.66	1.67
Guarulhos	4	2017-01-01	2024-01-01	53.08	3.71
Manaus	5	2025-01-01	2026-01-01	17.0	0.17
Osasco	15	2017-01-01	2024-01-01	51.72	3.62
Ribeirão Preto	10	2017-01-01	2024-01-01	49.1	3.44
Rio de Janeiro	20	2023-01-01	2026-01-01	61.4	1.84
Santo André	1	2017-01-01	2024-01-01	41.25	2.89
Sorocaba	3	2017-01-01	2024-01-01	52.92	3.7
São Bernardo do Campo	21	2017-01-01	2024-01-01	50.96	3.57
São Paulo	45	2017-01-01	2026-01-01	51.89	4.67

A etapa de coleta por meio da OpenAQ retornou dados efetivos para uma parcela dos 35 municípios inicialmente selecionados. Conforme resumido na Tabela 1 (informações dos sensores por cidade), foram identificadas 12 cidades com sensores e séries utilizáveis; o número de sensores por município variou consideravelmente, com maior densidade em áreas metropolitanas. Em particular, observou-se elevada concentração de sensores em São Paulo (45 sensores), seguido por São Bernardo do Campo (21), Rio de Janeiro (20) e Duque de Caxias (20). Cidades de porte menor ou com redes locais menos consolidadas exibiram contagens reduzidas (por exemplo, Sorocaba, Guarulhos e Manaus, com apenas alguns sensores cada). A Tabela 1 apresenta, para cada cidade, o total de sensores identificados, o período coberto pelas séries e a cobertura média anual (percentCoverage), indicadores fundamentais para avaliar a robustez das estimativas por município.

A agregação dos dados mensais por cidade permitiu estimar médias mensais e anuais por poluente e calcular o indicador de excedência frente às diretrizes da OMS. Os resultados de excedência estão consolidados na Tabela 2 (percentual de meses em que os níveis superaram os limites da OMS). Verificou-se que várias cidades apresentaram percentuais de meses com valores acima dos limites: por exemplo, Campinas e Guarulhos figuram entre as cidades com maior proporção de meses em excedência para PM_{2.5} (acima de 50% dos meses disponíveis), seguidas por Osasco, Ribeirão Preto e São Paulo com percentuais expressivos (na faixa de 35–45%). Ao mesmo tempo, algumas cidades não registraram excedência para determinados poluentes no conjunto de meses disponíveis; contudo, essa ausência não implica necessariamente conformidade contínua, pois muitas séries apresentam lacunas temporais importantes.

Em termos temporais, as séries mensais mostraram padrões sazonais coerentes com a literatura: concentrações de material particulado tendem a elevar-se em meses de menor pluviosidade e menor dispersão atmosférica, enquanto o ozônio tende a apresentar picos em períodos de maior radiação. Entretanto, a heterogeneidade da cobertura, tanto em número de sensores quanto em percentuais de cobertura anual entre cidades (variando entre valores muito baixos e moderados), limita a capacidade de generalizar tendências nacionais ou de comparar cidades em pé de igualdade. Por essa razão, os resultados apresentados nas Tabelas 1 e 2 devem ser interpretados tendo em conta as diferenças de densidade de monitoramento e de completude temporal entre os municípios.

4. Discussão

A análise dos dados coletados a partir da plataforma OpenAQ revelou padrões relevantes sobre a variação espacial, temporal e a incidência de excedências dos limites de qualidade do ar estabelecidos pela Organização Mundial da Saúde (OMS) nas grandes cidades brasileiras. Apesar de o estudo ter inicialmente contemplado 35 municípios com população superior a 500 mil habitantes, apenas 12 deles apresentaram sensores disponíveis e dados utilizáveis nos arquivos consolidados. Essa limitação de cobertura observacional deve ser enfatizada, pois influencia diretamente a capacidade de generalização dos resultados.

Do ponto de vista da infraestrutura de monitoramento, observou-se elevada disparidade entre as cidades. São Paulo, por exemplo, apresentou 45 sensores ativos, seguida de municípios como São Bernardo do Campo (21), Rio de Janeiro (20) e Duque de Caxias (20). Em contraste, cidades como Sorocaba, Guarulhos e Manaus apresentaram entre três

e cinco sensores apenas. Além disso, a cobertura temporal também se mostrou irregular, com percentuais médios de reporte variando significativamente — do mínimo aproximado de 17% em Manaus ao máximo de 61% no Rio de Janeiro. Essa heterogeneidade implica que algumas estimativas são mais robustas do que outras, especialmente em cidades com baixa cobertura ou poucos sensores, onde médias podem ser influenciadas por leituras atípicas.

No que se refere à variação espacial das concentrações médias anuais (RQ1), os resultados apontam que municípios da Região Metropolitana de São Paulo e do interior paulista concentram os maiores valores médios de poluentes. Para PM_{2.5}, destacam-se Campinas (~17,0 µg/m³), Ribeirão Preto (~16,8 µg/m³) e Guarulhos (~16,6 µg/m³), todos acima da média de São Paulo (~15,3 µg/m³). Para PM₁₀, Ribeirão Preto lidera com ~36,2 µg/m³, seguida pelo Rio de Janeiro (~31 µg/m³). No caso do NO₂, o Rio de Janeiro apresenta os maiores valores médios ~43,4 µg/m³, seguido por São Paulo (~38,7 µg/m³). Em geral, municípios altamente urbanizados e com maior fluxo veicular e industrial tendem a apresentar maiores concentrações, refletindo padrões coerentes com a literatura e com o comportamento esperado para áreas metropolitanas densas. Esses resultados reforçam a necessidade de políticas regionais integradas, especialmente no eixo da macrometrópole paulista.

Quanto aos padrões temporais e sazonais (RQ2), os dados mensais evidenciaram a presença de ciclos marcantes ao longo do ano. Os poluentes particulados (PM_{2.5} e PM₁₀) apresentaram pico sistemático nos meses de inverno (junho a agosto), período marcado por baixa dispersão atmosférica, condições estáveis e menor precipitação em boa parte do Sudeste e interior. Os valores médios de PM_{2.5}, por exemplo, atingiram cerca de 19–20 µg/m³ nesses meses. Por outro lado, o ozônio (O₃) apresentou picos nos meses de primavera e início do verão, comportamento esperado devido à intensificação da radiação solar e da dinâmica fotoquímica que rege sua formação. Esses padrões sazonais reforçam o entendimento científico consolidado sobre meteorologia e sua influência na dinâmica dos poluentes.

No tocante à excedência dos limites recomendados pela OMS (RQ3), observou-se percentuais elevados em diversos municípios quando se considera o número de meses disponíveis com médias acima do limite de referência. Para PM_{2.5}, cidades como Campinas (~53%), Guarulhos (~51%), Osasco (~45%) e Ribeirão Preto (~44%) apresentaram excedência em mais da metade dos meses monitorados. Mesmo São Paulo, com maior número de sensores e série histórica mais extensa, exibiu excedência de aproximadamente 37% dos meses. Cidades como Manaus, Belém e Rio de Janeiro não apresentaram excedências para esse poluente nos períodos observados; entretanto, essa ausência deve ser interpretada com cautela, pois em alguns casos os dados disponíveis concentram-se em períodos específicos, o que pode não refletir o comportamento anual típico. Assim, os resultados confirmam que, para uma parte significativa das cidades observadas, os níveis de PM_{2.5} ultrapassam com frequência os limites considerados seguros para a saúde humana.

Por fim, destaca-se que a capacidade de análise detalhada foi limitada pela irregularidade temporal, número desigual de sensores por cidade e pela ausência de dados em 23 das cidades inicialmente planejadas na pesquisa. A heterogeneidade da amostragem implica que algumas séries temporais oferecem alta confiabilidade, enquanto outras apre-

sentam variabilidade indesejada associada à escassez de dados. Ainda assim, os padrões identificados são consistentes com o conhecimento científico atual sobre qualidade do ar no Brasil e refletem preocupações reais, especialmente nas regiões mais urbanizadas.

References

- IBGE — Instituto Brasileiro de Geografia e Estatística (2008). Classificação e caracterização dos espaços urbanos e rurais do Brasil. Relatório Técnico. Define cidades de grande porte como aquelas com mais de 500 mil habitantes.
- Jornal, O. (2024). Fumaça de queimadas na Amazônia carrega partículas de poluição que afetam a saúde em todo o país. <https://oglobo.globo.com/>. Reportagem jornalística sobre impactos de queimadas e PM_{2.5}.
- Silva, J. and Santos, Maria e Almeida, J. (2021). Aumento da mortalidade por doenças crônicas não transmissíveis associado à poluição do ar no Brasil (2006–2016). *Cadernos do Leste*, 21(21):100–115.
- World Health Organization (2021). *WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*. WHO, Geneva.

Table 2. Tabela de Dados de Poluentes Excedentes por Cidade

Cidade	Poluente	Total de meses	Meses excedentes	Porcentagem de excedência
Belém	pm25	5	0	0.0
Campinas	co	32	0	0.0
Campinas	no2	24	1	4.17
Campinas	o3	33	0	0.0
Campinas	pm10	33	0	0.0
Campinas	pm25	30	16	53.33
Duque de Caxias	co	30	1	3.33
Duque de Caxias	no2	30	27	90.0
Duque de Caxias	o3	28	0	0.0
Duque de Caxias	pm10	31	2	6.45
Duque de Caxias	pm25	27	5	18.52
Duque de Caxias	so2	10	0	0.0
Guarulhos	no2	5	5	100.0
Guarulhos	o3	33	0	0.0
Guarulhos	pm10	33	4	12.12
Guarulhos	pm25	33	17	51.52
Manaus	pm25	4	0	0.0
Osasco	co	33	0	0.0
Osasco	no2	33	32	96.97
Osasco	o3	33	0	0.0
Osasco	pm10	33	5	15.15
Osasco	pm25	33	15	45.45
Osasco	so2	33	0	0.0
Ribeirão Preto	co	31	0	0.0
Ribeirão Preto	no2	32	1	3.12
Ribeirão Preto	o3	30	0	0.0
Ribeirão Preto	pm10	32	9	28.12
Ribeirão Preto	pm25	32	14	43.75
Rio de Janeiro	co	27	0	0.0
Rio de Janeiro	no2	30	26	86.67
Rio de Janeiro	o3	31	0	0.0
Rio de Janeiro	pm10	31	4	12.9
Rio de Janeiro	pm25	4	0	0.0
Santo André	o3	26	0	0.0
Sorocaba	no2	33	4	12.12
Sorocaba	o3	33	0	0.0
Sorocaba	pm10	33	2	6.06
São Bernardo do Campo	co	31	0	0.0
São Bernardo do Campo	no2	33	16	48.48
São Bernardo do Campo	o3	33	0	0.0
São Bernardo do Campo	pm10	33	2	6.06
São Bernardo do Campo	pm25	31	11	35.48
São Bernardo do Campo	so2	33	0	0.0
São Paulo	co	33	0	0.0
São Paulo	no2	33	32	96.97
São Paulo	o3	33	0	0.0
São Paulo	pm10	33	2	6.06
São Paulo	pm25	56	21	37.5
São Paulo	so2	33	0	0.0