

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

Jhonata Silveira Dias<sup>1</sup>, Lucca Oliveira Vasconcelos de Faria<sup>1</sup>,  
Pedro Henrique Moreira Caixeta Ferreira<sup>1</sup>

<sup>1</sup>Instituto 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, lovfaria}@sga.pucminas.br, phmcf13@gmail.com

**Abstract.** *This meta-paper describes the style to be used in articles and short papers for SBC conferences. For papers in English, you should add just an abstract while for the papers in Portuguese, we also ask for an abstract in Portuguese (“resumo”). In both cases, abstracts should not have more than 10 lines and must be in the first page of the paper.*

**Resumo.** *Este meta-artigo descreve o estilo a ser usado na confecção de artigos e resumos de artigos para publicação nos anais das conferências organizadas pela SBC. É solicitada a escrita de resumo e abstract apenas para os artigos escritos em português. Artigos em inglês deverão apresentar apenas abstract. Nos dois casos, o autor deve tomar cuidado para que o resumo (e o abstract) não ultrapassem 10 linhas cada, sendo que ambos devem estar na primeira página do artigo.*

## 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<sub>2.5</sub>), o material particulado inalável (PM<sub>10</sub>), SO<sub>2</sub>, NO<sub>2</sub>, CO e O<sub>3</sub>.

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. 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. 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. 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. Goal

Investigar os padrões espaciais e temporais da qualidade do ar nas cidades brasileiras, identificando tendências, possíveis fatores de influência e comparando os níveis observados com os limites recomendados pela OMS.

### 2.2. Questions

- 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

### 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, SO<sub>2</sub>, NO<sub>2</sub>, CO e O<sub>3</sub> 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 de coleta em si pode ser separado em duas etapas principais. A primeira foi a coleta dos sensores para cada cidade. Em seguida, para cada sensor foram reunidos seus dados mensais e anuais.

## 3. Resultados

In some conferences, the papers are published on CD-ROM while only the abstract is published in the printed Proceedings. In this case, authors are invited to prepare two final versions of the paper. One, complete, to be published on the CD and the other, containing only the first page, with abstract and “resumo” (for papers in Portuguese).

## 4. Discussão

Section titles must be in boldface, 13pt, flush left. There should be an extra 12 pt of space before each title. Section numbering is optional. The first paragraph of each section should not be indented, while the first lines of subsequent paragraphs should be indented by 1.27 cm.

## References