ANDREI SOUZA DE OLIVEIRA

GABRIELA OHASHI DE SOUZA

MARINA OHASHI DE SOUZA

MATHEUS FERREIRA FRANÇA

MIGUEL MAURÍCIO TADEU PITALI DA SILVA

**PROJETO APLICADO I**

**Poluição atmosférica no estado de São Paulo e efeitos nocivos na saúde da população, no período de 2015 a 2021**

São Paulo

2023

**SUMÁRIO**

1. INTRODUÇÃO ......................................................................................... 3
2. PREMISSAS ............................................................................................ 4
3. OBJETIVOS E METAS ............................................................................ 5
4. GITHUB DO PROJETO ........................................................................... 6
5. CONTEXTO DO ESTUDO ....................................................................... 7
6. REFERÊNCIAS DE AQUISIÇÃO DO DATASET ..................................... 8
7. DESCRIÇÃO DA ORIGEM ...................................................................... 9
8. DESCRIÇÃO DO DATASET ................................................................... 10
9. ANÁLISE EXPLORATÓRIA DOS DADOS...............................................11
10. LISTA DE TABELAS ............................................................................... 18
11. CRONOGRAMA DO PROJETO ............................................................. 19

**INTRODUÇÃO**

A poluição ambiental vem causando impactos, nos últimos anos, e dados preocupantes sobre eles nos levam a refletir sobre o efeito deles em nossas vidas. O Brasil, que passou por um grande processo de urbanização, tende a sofrer ainda mais com as consequências desses impactos, e em uma cidade como São Paulo essa premissa não seria diferente. Os problemas de saúde são apenas uma pequena parcela dessas consequências.

**PREMISSAS**

Avaliar a poluição atmosférica no estado de São Paulo, no período de 2015 a 2021, e relacioná-la aos efeitos nocivos na saúde da população, com base no índice brasileiro da qualidade do ar definido pelo IEMA (Instituto de Energia e Meio Ambiente).

Utilização de base de dados disponível no Kaggle: <https://www.kaggle.com/datasets/samirnunesdasilva/sao-paulo-pollution-data>

**OBJETIVOS E METAS**

Identificar:

* Poluente mais crítico
* Localidade mais afetada pelo CO
* Período do ano com maior concentração de CO
* Período do ano com menor concentração de CO
* Horário com maior concentração de CO

Correlacionando esses indicadores com os efeitos na saúde da população.

Tabela 1 – Estrutura do índice brasileiro de qualidade do ar e efeitos à saúde

Tabela

Descrição gerada automaticamente

**GITHUB DO PROJETO**

[https://github.com/OhashiMarina/Projeto-Aplicado-I.git](https://bra01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgithub.com%2FOhashiMarina%2FProjeto-Aplicado-I.git&data=05%7C01%7C%7Ca159a47157094da4aa8808db22ff7cec%7C84df9e7fe9f640afb435aaaaaaaaaaaa%7C1%7C0%7C638142252929122044%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=ddUPTaWCIXU%2BMwoq8OggR21JD%2BVL1BKK6QvFdLzdnuE%3D&reserved=0)

**CONTEXTO DO ESTUDO**

Avaliar a poluição atmosférica no estado de São Paulo, no período de 2015 a 2021, e relacioná-la aos efeitos nocivos na saúde da população, com base no índice brasileiro da qualidade do ar definido pelo IEMA (Instituto de Energia e Meio Ambiente).

**REFERÊNCIAS DE AQUISIÇÃO DO DATASET**

Utilização de base de dados disponível no Kaggle - [https://www.kaggle.com/datasets/samirnunesdasilva/sao-paulo-pollution-data](https://bra01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.kaggle.com%2Fdatasets%2Fsamirnunesdasilva%2Fsao-paulo-pollution-data&data=05%7C01%7C%7C8745d2d1f14c4ab6787008db22f3c52b%7C84df9e7fe9f640afb435aaaaaaaaaaaa%7C1%7C0%7C638142202605758390%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=twlb4MXFmjhcUAs5%2Ft8TtM2WDPAX6A8DX%2FqVO7UEvLA%3D&reserved=0)

Período: 2015 a 2021

Conjunto de dados com mais de 10 milhões de linhas englobando 87 estações de medição distintas distribuídas ao longo do estado de São Paulo que podem medir a concentração de 9 poluentes: MP10, O3, NO2, MP2.5, CO, SO2, NO, FMC e PTS.

**DESCRIÇÃO DA ORIGEM**

Fonte: Plataforma Qualidade do Ar, do Instituto de Energia e Meio Ambiente (IEMA);  
Site: <https://energiaeambiente.org.br/qualidadedoar#secao-14>.

**DESCRIÇÃO DO DATASET**

Colunas possuem as seguintes descrições:

**ID:** identificação, via índice inteiro, de cada registro do dataframe (chave primária);

**Data:** data em que foi feita a medição da concentração do poluente;

**Hora:** hora em que foi feita a medição;

**Estação:** local em que foi feita a medição;

**Código:** código associado à estação em que foi realizada a medição;

**Poluente:** poluente cuja concentração foi medida;

**Valor:** valor, na unidade especificada, da concentração do poluente;

**Unidade:** unidade de concentração utilizada; e

**Tipo:** como foi realizada a medição (de forma automática ou manual).

**ANÁLISE EXPLORATÓRIA DOS DADOS**

Realizamos os seguintes procedimentos em Python para a exploração dos dados:

**Avaliação primária dos dados:**

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

**Condicional dos Aferíveis**

Considerando a necessidade de correlacionar os indicadores estabelecidos no objetivo do projeto com a Tabela 1, identificou-se a necessidade de exclusão dos poluentes PTS, FMC e NO da base de dados.

Texto

Descrição gerada automaticamente

**Criação de novo *dataframe***

Novo *dataframe* considerando apenas os poluentes a serem analisados, conforme Tabela 1.

Interface gráfica do usuário, Tabela

Descrição gerada automaticamente

**Respondendo às perguntas do Objetivo do Projeto**

**1 - Poluente mais crítico**

Foi necessário definir a regra de classificação considerando a Tabela 1

Calendário

Descrição gerada automaticamente com confiança média

Foi Atribuído um índice de nocividade de 0 a 5, sendo 0 o menos nocivo e 5 o mais nocivo, podemos notar que o poluente CO apresenta um maior risco à saúde que os demais poluentes.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Texto

Descrição gerada automaticamente

**2 - Período com maior nível de concentração de CO**

Foi criada uma nova coluna agrupando as medições por mês/ano e identificados os 10 períodos de maior concentração de CO.

Texto

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Tabela

Descrição gerada automaticamente

Gráfico, Gráfico de barras

Descrição gerada automaticamente

É possível notar que os 10 piores períodos ocorrem no inverno, sendo 8 nos meses de junho e julho. Época em que a umidade relativa do ar é menor.

**3 – Período com menor concentração de CO**

Identificados os 10 períodos de menor concentração de CO.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamenteTexto, Tabela

Descrição gerada automaticamenteInterface gráfica do usuário, Texto, Email

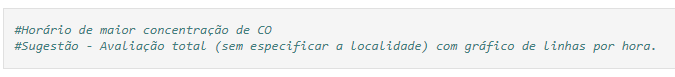
Descrição gerada automaticamenteGráfico, Gráfico de barras

Descrição gerada automaticamente

Os períodos de menor concentração de CO ocorrem na primavera e no verão, o que pode estar relacionado com a maior frequência de chuvas.

**4 - Horário de maior concentração de CO**

As medições foram agrupadas por hora para identificação dos horários de maior concentração de CO.



Uma imagem contendo Texto

Descrição gerada automaticamenteTexto

Descrição gerada automaticamenteTabela

Descrição gerada automaticamenteGráfico, Gráfico de linhas

Descrição gerada automaticamente

Podemos observar que os horários de maior concentração de CO ocorrem no período da manhã (entre 08h e 09h) e no final da tarde (entre 19h e 23h).

Tal fato pode estar relacionado aos horários de pico em que há um maior volume de carros em circulação.

**LISTA DE TABELAS**

Tabela 1 – Estrutura do índice brasileiro de qualidade do ar e efeitos à saúde

Tabela

Descrição gerada automaticamente

**CRONOGRAMA DO PROJETO**

Tabela

Descrição gerada automaticamente com confiança média

Arquivo Excel disponível no Github.

Observação: Dias restantes contados da data da elaboração desse documento, 27/05/2023.