

Vinícius Kurek  
Eduardo Kurek

## **Etapa 1**

### **Trabalho Final**

Relatório técnico parcial do trabalho final solicitado pelo professor Eduardo Pena na disciplina de Ciência de Dados do Bacharelado em Ciência da Computação da Universidade Tecnológica Federal do Paraná.

Universidade Tecnológica Federal do Paraná – UTFPR

Departamento Acadêmico de Computação – DACOM

Bacharelado em Ciência da Computação – BCC

Campo Mourão

Outubro / 2025

## 1 Descrição

O *dataset* escolhido foi o [Air Quality NYC](#). São dados que possuem os índices de gases emitidos na cidade de NY e suas principais fontes. Também possui dados hospitalares causados por esses poluentes, tráfegos de caminhões e carros. O tamanho inicial da tabela é de 18.862 entradas.

## 2 Justificação

Analisar os dados relacionados à qualidade do ar é de extrema importância, pois com base nesses dados, medidas governamentais e judiciárias podem ser tomadas para diminuir o impacto causado por esses poluentes. Pesquisas por atividades mais sustentáveis podem ser iniciadas a fim de melhorar a qualidade de vida da população.

## 3 Perguntas e Hipóteses

- **P1:** É possível prever a taxa de poluição emitida no ar nos próximos anos?
- **P2:** A estação do ano influencia a quantidade de poluentes no ar de alguma forma?
- **H1:** Há uma correlação negativa significativa entre o período do ano e a poluição do ar.
- **H2:** O período passado é um bom preditor para dizer se a poluição do ar vai subir ou diminuir no próximo ano.
- **H3:** Gases emitidos por combustão em caldeira e tráfego de caminhões são consideravelmente mais prejudiciais do que gases emitidos pelo tráfego de carros.

## 4 Github

<https://github.com/ViniciusKurek/Projeto-Final—Data-Science>