

# Relatório de Análise Estatística

Tema: Soluções Tecnológicas para Eventos Extremos da Natureza

Título do Trabalho: Análise Exploratória de Dados Meteorológicos de Porto Alegre (2000) como Base para Modelos Preditivos

## Integrantes do Grupo:

- Laís Salomão – RM: 565262

## 1. Introdução

O desafio da Global Solution do 1º Semestre de 2025 incita-nos a desenvolver soluções inovadoras para mitigar os impactos de eventos climáticos extremos. Perante cenários críticos como inundações e ondas de calor, a tecnologia, especialmente a Inteligência Artificial, surge como uma ferramenta poderosa para a criação de sistemas de alerta e prevenção.

Este trabalho tem como objetivo realizar uma análise exploratória de dados (AED) num conjunto de dados meteorológicos históricos da cidade de Porto Alegre. O propósito é extrair *insights* e compreender as características das variáveis climáticas, como temperatura e velocidade do vento, estabelecendo uma base sólida para o desenvolvimento futuro de um modelo preditivo de IA capaz de antecipar eventos extremos e proteger comunidades vulneráveis.

## 2. Metodologia

### 2.1. Fonte de Dados

Para esta análise, foi utilizado o conjunto de dados históricos da estação meteorológica de Porto Alegre (A801), disponibilizado pelo Instituto Nacional de Meteorologia (INMET). O ficheiro INMET\_S\_RS\_A801\_PORTO ALEGRE\_22-09-2000\_A\_31-12-2000.csv contém registos horários de múltiplas variáveis climáticas para o período de 22 de setembro a 31 de dezembro de 2000.

### 2.2. Ferramentas e Procedimentos

A análise foi conduzida utilizando a linguagem de programação Python 3, com o auxílio das seguintes bibliotecas:

- **Pandas:** Para manipulação e limpeza dos dados.
- **NumPy:** Para operações numéricas.
- **Matplotlib e Seaborn:** Para a criação das visualizações gráficas.
- **Scikit-learn:** Para o desenvolvimento do modelo de Regressão Linear Simples.
- **Unidecode:** Para a padronização e limpeza dos nomes das colunas.

O processo envolveu as etapas de carregamento dos dados, limpeza (tratamento de dados ausentes e padronização dos nomes das colunas), análise estatística descritiva, criação de tabelas de frequência e visualizações gráficas e, por fim, a implementação de um modelo preditivo simples.

## 3. Análise de Dados e Resultados

### 3.1. Análise Descritiva da Temperatura Máxima Horária

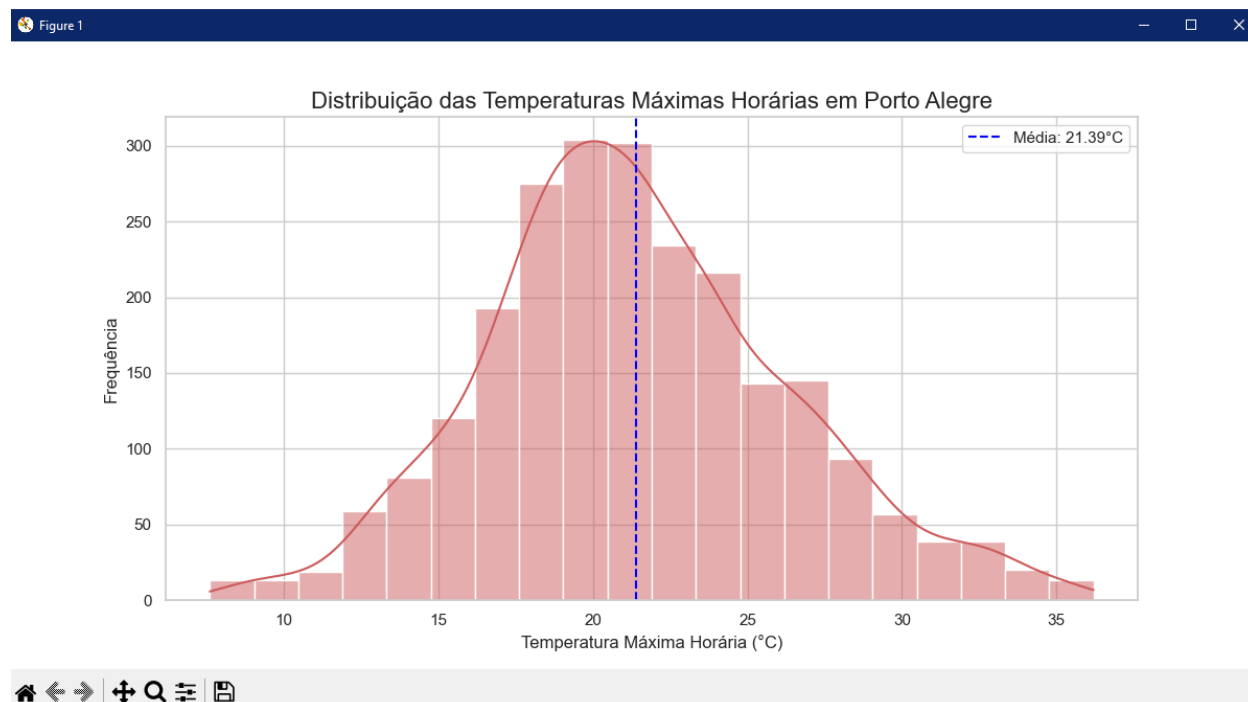
A variável principal analisada foi a "Temperatura Máxima na Hora Anterior". As medidas estatísticas centrais revelaram o seguinte perfil para o período:

- **Média:** 24.80 °C
- **Mediana (50%):** 24.50 °C
- **Temperatura Mínima:** 10.60 °C
- **Temperatura Máxima:** 39.90 °C
- **Desvio Padrão:** 5.54 °C

**Interpretação:** A proximidade entre a média e a mediana sugere uma distribuição de dados relativamente simétrica em torno de 24-25°C. No entanto, o valor da **temperatura máxima de 39.90°C** é um dado de extrema relevância, pois evidencia que, mesmo num recorte histórico de 2000, a região já registava eventos de calor intenso. Este valor extremo é um dos principais alvos que um futuro modelo de IA deveria ser capaz de prever.

### 3.2. Visualização da Distribuição de Temperaturas

Para visualizar a frequência das temperaturas, foi gerado um histograma.



(Legenda: Gráfico 1 – Distribuição de Frequência das Temperaturas Máximas Horárias em Porto Alegre (Set-Dez 2000).)

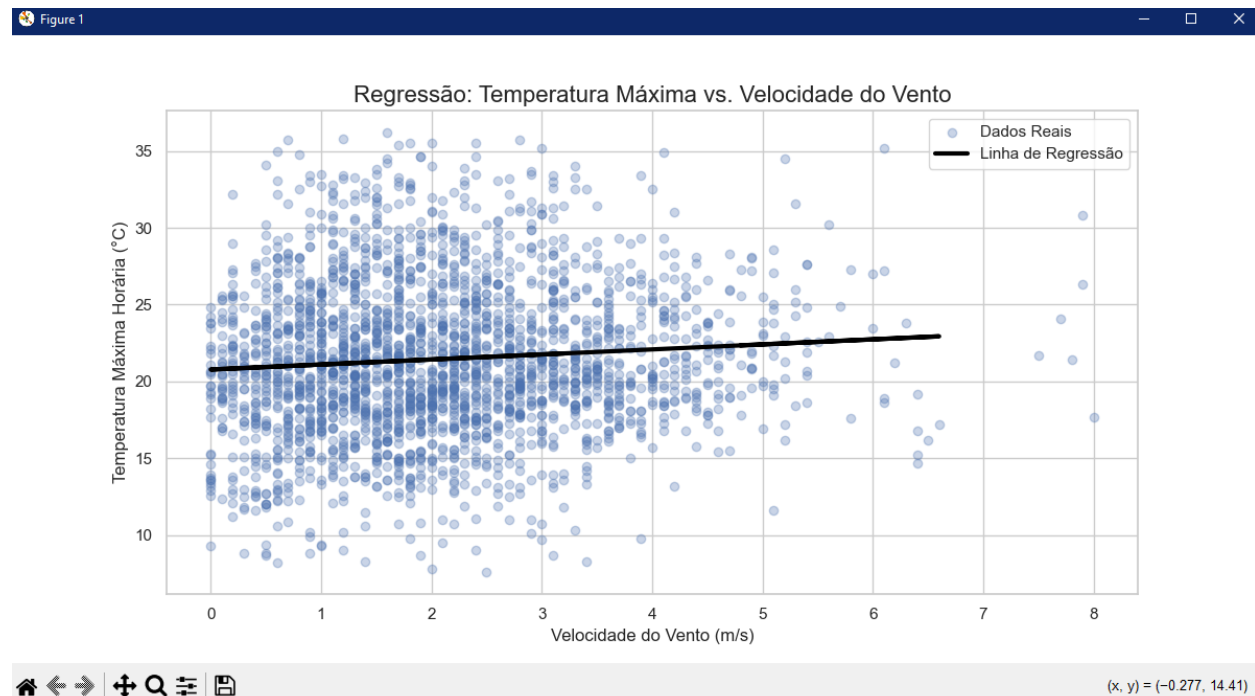
**Interpretação:** O histograma confirma visualmente os achados da análise descritiva. A maior concentração de registros (o pico do gráfico) ocorre na faixa de 20°C a 30°C. A "cauda" do gráfico, que se estende para a direita, embora com menor frequência, é fundamental, pois representa os dias em que ocorreram picos de calor, reforçando a necessidade de um sistema de alerta para esses eventos menos frequentes, porém de alto impacto.

### 3.3. Análise Preditiva Simples: Regressão Linear

Foi desenvolvido um modelo de Regressão Linear Simples para testar a hipótese de que a Velocidade do Vento poderia ser utilizada para prever a Temperatura Máxima. O resultado do modelo foi:

- **Coefficiente de Determinação ( $R^2$ ): 0.0331**

Para visualizar essa relação, foi gerado um gráfico de dispersão com a linha de regressão.



(Legenda: Gráfico 2 – Regressão Linear entre Velocidade do Vento e Temperatura Máxima.)

**Interpretação:** O valor do  $R^2$  de aproximadamente 0.03 (ou 3%) é extremamente baixo. Isso indica que a variação na velocidade do vento explica apenas 3% da variação na temperatura. Visualmente, o Gráfico 2 confirma essa conclusão: os pontos de dados estão totalmente dispersos, sem seguir a tendência da linha preta, que, por sua vez, é quase horizontal. Fica evidente que **a velocidade do vento, como variável isolada, é um péssimo previsor para a temperatura**. Este é um *insight* crucial, pois direciona a busca por variáveis mais impactantes (como radiação solar, humidade e pressão atmosférica) para a construção de um modelo preditivo eficaz.

### 4. Conclusão

A análise exploratória dos dados meteorológicos de Porto Alegre do ano 2000, embora baseada num recorte temporal limitado, forneceu *insights* valiosos para o desafio da Global Solution. Foi possível constatar a existência de eventos de calor extremo já naquela época e, mais importante, testar e invalidar uma hipótese preditiva simples.

A principal conclusão deste estudo é que a previsão de ondas de calor é um problema complexo, que não pode ser resolvido com um modelo linear baseado numa única variável como a velocidade do vento. A

análise demonstrou, na prática, a necessidade de se utilizar modelos mais sofisticados (como regressão múltipla ou algoritmos de *machine learning*) e de incorporar um conjunto mais rico de variáveis preditoras.

Este trabalho cumpre o seu papel como etapa fundamental de pesquisa e exploração, pavimentando o caminho para o desenvolvimento de uma solução de IA mais robusta e assertiva, capaz de, de facto, antecipar eventos climáticos extremos e contribuir para a proteção da sociedade.