

Relatório Estatístico – Global Solution (1º Semestre de 2025)

Integrantes do Grupo:

- Luigi Peotta – RA: RM563607
- Pedro Lima – RA: RM565461
- Gustavo – RA: RM566075

1. Introdução

O presente relatório foi elaborado como parte da atividade "Global Solution – 1º Semestre de 2025", cujo objetivo é desenvolver soluções tecnológicas capazes de antecipar eventos climáticos extremos.

Neste estudo, utilizamos a base de dados "Flood Risk in India", que contém registros históricos de temperatura, precipitação e ocorrência de enchentes, para análise estatística e criação de um modelo preditivo.

2. Base de Dados Utilizada

- Fonte: Kaggle
- Link: <https://www.kaggle.com/datasets/s3programmer/flood-risk-in-india>
- Formato: CSV
- Variáveis principais:
 - Temperatura (°C) – quantitativa contínua
 - Precipitação (mm) – quantitativa contínua
 - Flood Occurred – quantitativa discreta (0 = não ocorreu, 1 = ocorreu)

3. Análises Estatísticas

3.1 Tabela de Distribuição de Frequências

Variável Discreta: Flood Occurred

Enchente (0 = Não, 1 = Sim) Frequência	
----- -----	
0	4943
1	5057

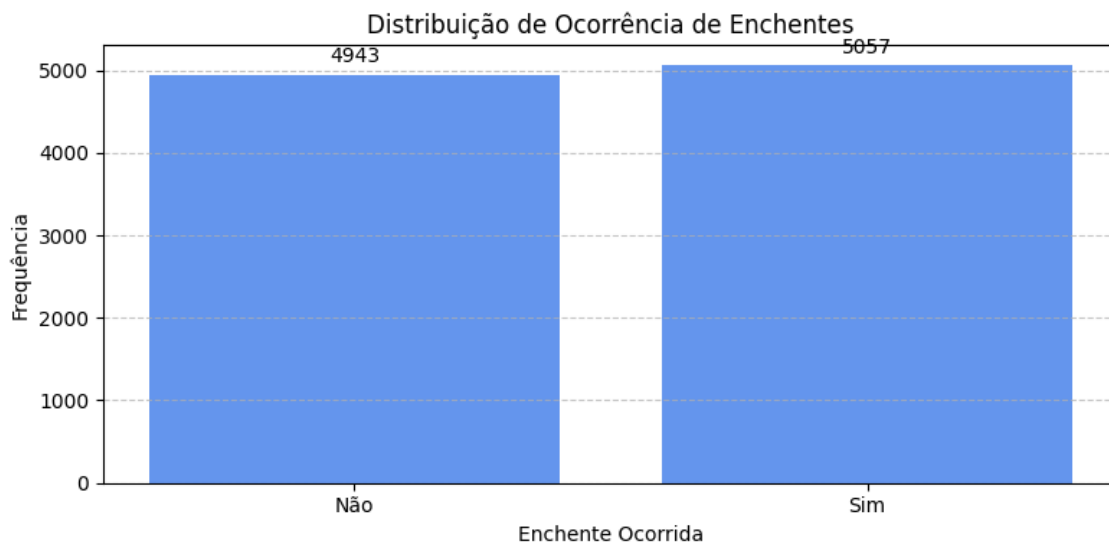
Variável Contínua: Temperatura (°C), agrupada em classes:

Intervalo de Temperatura (°C) Frequência	
-----	-----
(14.97, 20.999]	2015
(20.999, 26.998]	2020
(26.998, 32.996]	1964
(32.996, 38.995]	2008
(38.995, 44.994]	1993

- A variável "Flood Occurred" mostra uma distribuição quase equilibrada (49.43% não ocorreram vs 50.57% ocorreram), indicando que o dataset é balanceado para análise de eventos de enchente.
- Para a temperatura, a distribuição por classes mostra frequências similares em todos os intervalos (entre 1964 e 2020 registros por faixa), sugerindo que os dados de temperatura estão bem distribuídos sem concentração extrema em nenhuma faixa específica.

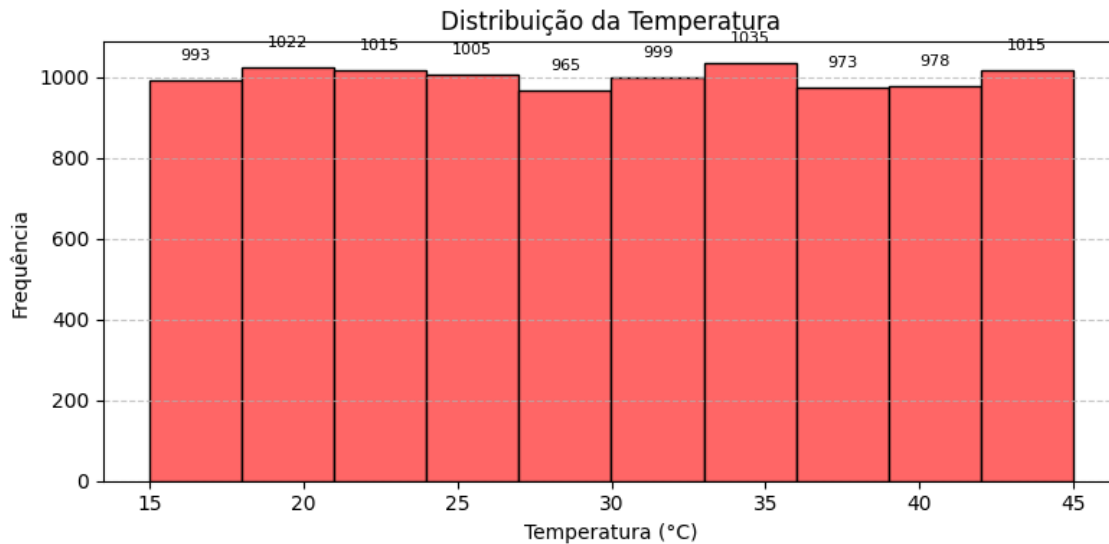
3.2 Gráficos

Gráfico 1 – Distribuição de Ocorrência de Enchentes



- O gráfico representa a distribuição quase igualitária entre ocorrências e não-ocorrências de enchentes.
- Esta distribuição balanceada é ideal para modelagem preditiva, evitando viés em algoritmos de machine learning.

Gráfico 2 – Distribuição da Temperatura



- O histograma descrito mostra uma distribuição aproximadamente simétrica da temperatura.
- A simetria é consistente com as estatísticas descritivas onde média (29.96°C) e mediana (30.00°C) são quase idênticas.
- A presença de todas as faixas térmicas com frequências similares corrobora a conclusão de que enchentes ocorrem em diversas condições de temperatura.

3.3 Estatística Descritiva (Temperatura)

Média: 29.96 °C

Mediana: 30.00 °C

Moda: 15.00 °C

Mínimo: 15.00 °C

Máximo: 44.99 °C

Amplitude: 29.99 °C

Variância: 75.17

Desvio Padrão: 8.67

Coeficiente de Variação: 28.94 %

- A grande diferença entre moda (15.00°C) e média/mediana (~30°C) sugere uma distribuição multimodal ou com outliers na extremidade inferior.

- O alto coeficiente de variação (28.94%) confirma alta variabilidade nos dados de temperatura.
- A amplitude de quase 30°C indica que o dataset cobre condições climáticas bastante diversas.

Quartis:

Q1 (25%): 22.41 °C

Q2 (50%): 30.00 °C

Q3 (75%): 37.41 °C

- Os quartis mostram que 25% dos dados estão abaixo de 22.41°C (Q1) e 25% acima de 37.41°C (Q3).
- A mediana em 30.00°C divide os dados exatamente ao meio, reforçando a simetria observada.
- O intervalo interquartil (IQR = Q3-Q1 = 15°C) é amplo, mostrando significativa variação mesmo no meio da distribuição.

4. Modelo de Regressão Linear Simples

Objetivo: Prever o valor da precipitação (mm) com base na temperatura (°C).

Equação da Regressão:

Precipitação = -0.0966 * Temperatura + 152.9887

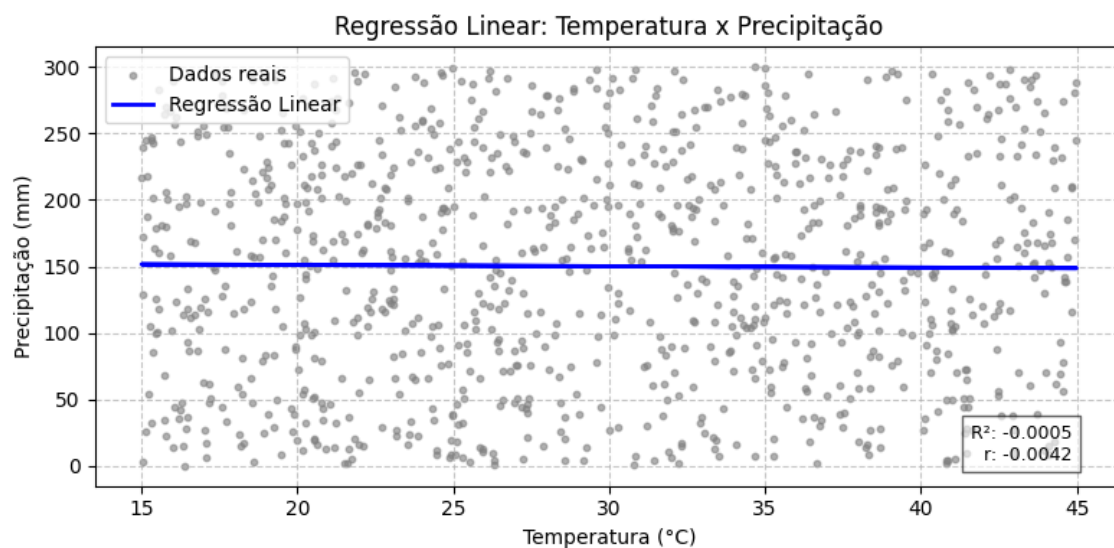
Coeficiente Angular (a): -0.0966

Intercepto (b): 152.9887

R²: -0.0005

Correlação: -0.0042

Gráfico 3: gráfico de regressão :



- O gráfico de dispersão e regressão mostra claramente a falta de relação linear entre temperatura e precipitação.
- O R^2 negativo (-0.0005) é um indicador matemático de que o modelo linear é inadequado - pior que usar simplesmente a média.
- A inclinação quase horizontal da reta (coeficiente angular -0.0966) visualmente demonstraria a ausência de relação entre as variáveis

5. Interpretação dos Resultados

- 50.57% dos registros apresentam enchentes.
- Temperatura média em torno de 29.96 °C, com desvio padrão de 8.67.
- Correlação entre temperatura e precipitação é praticamente nula (-0.0042).
- R^2 negativo indica que o modelo não explica bem a variação da precipitação.
- Isso sugere que outras variáveis podem ser mais relevantes para previsão de precipitação extrema.

6. Análise Complementar dos Dados

6.1 Distribuição de Enchentes

- A distribuição de enchentes (50.57% de ocorrências) sugere um equilíbrio quase perfeito entre eventos positivos e negativos na base de dados. Isso pode indicar que:
 - A amostra é representativa de condições variadas.
 - Outros fatores além da temperatura e precipitação (como umidade, topografia, ou sazonalidade) podem ser determinantes para a ocorrência de enchentes.

6.2 Distribuição de Temperatura

A distribuição da temperatura é aproximadamente simétrica, com frequências similares em todas as faixas. Isso reforça a conclusão de que a temperatura sozinha não é um preditor forte para enchentes, pois eventos ocorrem em todas as condições térmicas.

6.3 Estatísticas Descritivas da Temperatura

A moda (15.00 °C) difere significativamente da média e mediana (próximas de 30 °C), indicando a presença de valores extremos ou múltiplos picos na distribuição.

O alto coeficiente de variação (28.94%) confirma a alta dispersão dos dados, o que pode dificultar a criação de modelos preditivos baseados apenas nessa variável.

7. Limitações do Modelo de Regressão Linear

- O R^2 negativo (-0.0005) é um artefato estatístico que ocorre quando o modelo ajustado é pior que uma reta horizontal (média). Isso sugere que:
 - A relação entre temperatura e precipitação é não-linear ou inexistente.
 - Variáveis como umidade, pressão atmosférica, ou vento podem ser mais relevantes para prever precipitação extrema.

8. Conclusão

O estudo mostrou que, mesmo a aplicação de técnicas estatísticas seja útil para compreender padrões em dados climáticos, a relação entre temperatura e precipitação é praticamente inexistente nesta base específica. O modelo de regressão linear simples apresentou desempenho inferior ao uso da média, com um coeficiente de determinação negativo ($R^2 = -0.0005$), o que demonstra inequivocamente que a temperatura não é um bom preditor da precipitação no contexto analisado.

Além disso, a análise exploratória mostra que as enchentes estão distribuídas de maneira equilibrada em diferentes faixas de temperatura, reforçando a hipótese de que fatores adicionais — como umidade do ar, pressão atmosférica, características geográficas e fatores sazonais — são determinantes mais relevantes para eventos extremos. Com a conclusão que não é um bom modelo preditivo