

FACULDADE DE INFORMÁTICA E ADMINISTRAÇÃO PAULISTA
CURSO DE GRADUAÇÃO TECNOLOGO EM INTELIGÊNCIA
ARTIFICIAL

AUGUSTO OLIVEIRA CODO – RM562080
FELIPE DE OLIVEIRA CABRAL – RM561720
SOFIA BUERIS NETTO – RM565818

GLOBAL SOLUTION — ANÁLISE ESTATÍSTICA PARA
MONITORAMENTO DE ENCHENTES URBANAS

SÃO PAULO, SP
2025

INTRODUÇÃO

Este relatório tem como objetivo apresentar a análise estatística descritiva aplicada sobre a base de dados "Flood Risk in India", obtida na plataforma Kaggle. A análise faz parte do desenvolvimento de uma Prova de Conceito (POC) para um sistema inteligente capaz de monitorar, prever e auxiliar na resposta rápida a enchentes urbanas, conforme proposto no Global Solution – 1º Semestre de 2025.

Por meio desta análise, busca-se extrair insights relevantes sobre variáveis ambientais, especialmente relacionadas à ocorrência de enchentes e volume de precipitação, permitindo fundamentar tecnicamente o desenvolvimento de soluções tecnológicas eficazes no enfrentamento de desastres naturais.

1. TABELA DE DISTRIBUIÇÃO DE FREQUÊNCIA

1.1 Variável Quantitativa Discreta — Historical Floods

Ocorrência de Enchente Histórica	Frequência
0 – Não ocorreu	5013
1 – Ocorreu	4987

(Tabela de distribuição de frequência da variável Historical Floods)

A tabela acima apresenta a frequência da variável Historical Floods, que indica se houve (1) ou não (0) registros de enchentes históricas nas localidades analisadas. Observa-se que os dados estão relativamente balanceados, com 5013 registros de localidades sem histórico de enchentes e 4987 registros de localidades que já enfrentaram enchentes no passado. Essa distribuição praticamente equitativa é relevante, pois demonstra que o conjunto de dados possui boa representatividade tanto de cenários com quanto sem ocorrência de eventos críticos. Essa característica é essencial para a construção de modelos preditivos mais robustos e equilibrados, permitindo que o sistema de monitoramento tenha uma capacidade eficiente de aprendizado e generalização no reconhecimento de padrões associados às enchentes.

1.2 Variável Quantitativa Contínua — Rainfall (mm)

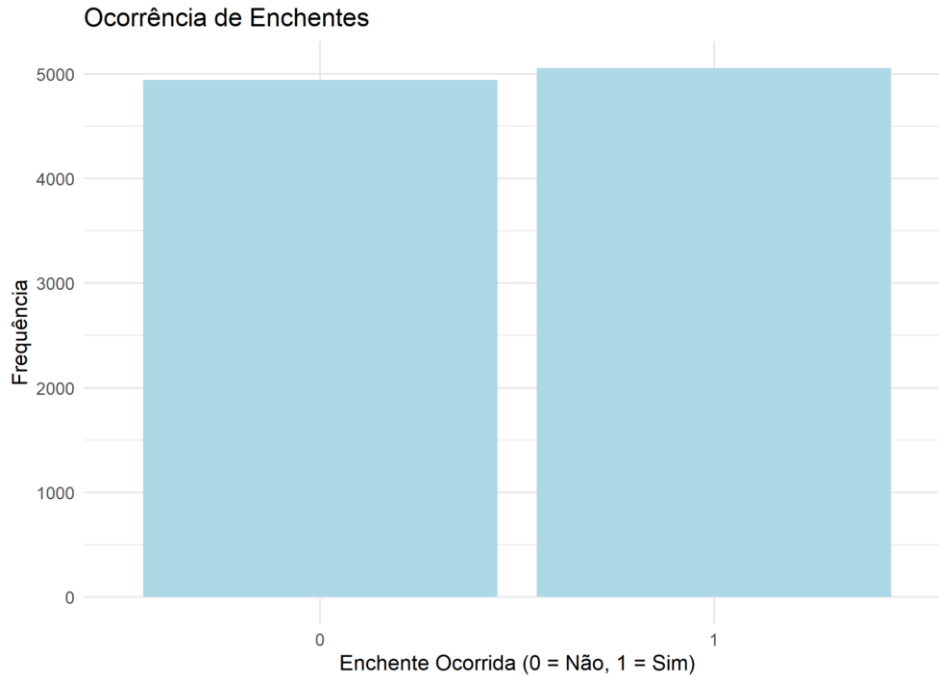
Intervalo de Chuva (mm)	Frequência
[-0.29, 50)	1635
[50, 100)	1685
[100, 150)	1658
[150, 200)	1734
[200, 250)	1669
[250, 300)	1619

(Tabela de distribuição de frequência da variável Rainfall (mm))

A variável Rainfall (mm) representa o volume de precipitação acumulada em cada localidade, mensurado em milímetros. A análise da distribuição dos dados, organizada em classes de amplitude regular, revela uma concentração relativamente uniforme entre os intervalos analisados. Nota-se, no entanto, uma leve predominância no intervalo de 150 mm a 200 mm, indicando que esse volume de chuva é ligeiramente mais recorrente nas amostras observadas. Essa distribuição balanceada é relevante, pois demonstra que há variação suficiente nos dados para alimentar modelos estatísticos e preditivos. Além disso, compreender essa distribuição é essencial para identificar os limiares críticos de precipitação que podem estar associados ao aumento do risco de enchentes nas regiões analisadas.

2. ANÁLISE GRÁFICA

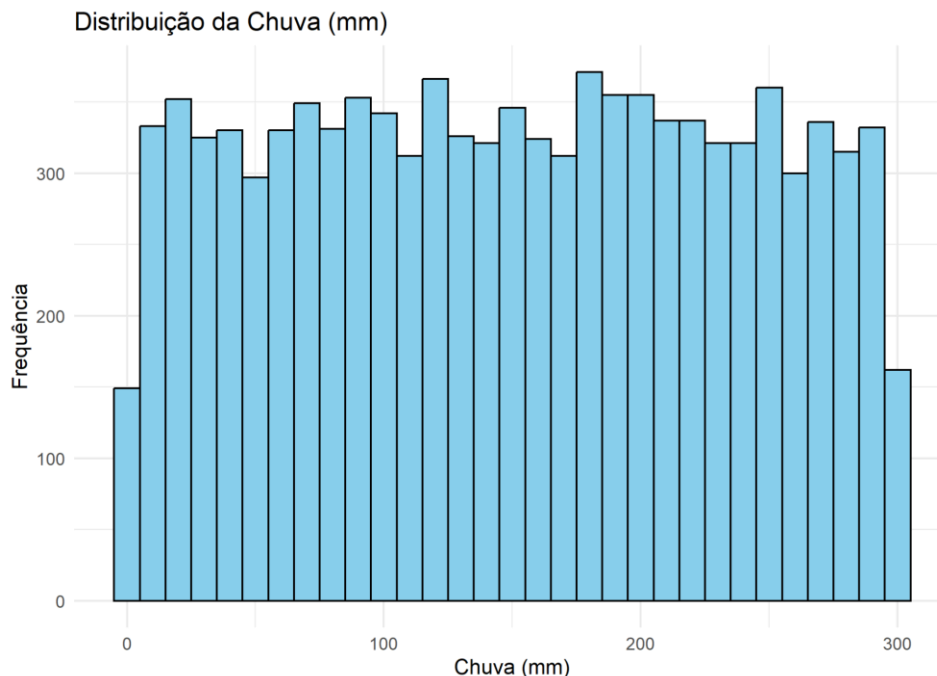
2.1 Gráfico de Barras — Flood Occurred



(Gráfico de barras representando a ocorrência de enchentes — Flood Occurred)

O gráfico de barras ilustra de forma clara a quantidade de localidades que registraram (1) ou não registraram (0) ocorrência de enchentes. Visualmente, observa-se um equilíbrio proporcional entre os dois grupos, o que indica que a base de dados possui uma boa representatividade tanto de cenários sem enchentes quanto daqueles em que o evento ocorreu. Essa distribuição balanceada é essencial para garantir que análises estatísticas e modelos preditivos desenvolvidos a partir desses dados não sofram viés, além de favorecer uma maior acurácia na identificação de padrões relacionados ao risco de enchentes.

2.2 Histograma — Rainfall (mm)



(Histograma representando a distribuição dos volumes de chuva — Rainfall (mm))

O histograma ilustra a distribuição dos volumes de precipitação (Rainfall (mm)) nas localidades analisadas. Observa-se que a maior concentração dos dados ocorre no intervalo de 50 mm a 250 mm, com destaque para um pico mais acentuado entre 150 mm e 200 mm, indicando que esse é o intervalo de chuva mais recorrente no conjunto de dados. Essa distribuição evidencia uma predominância de volumes médios a altos de precipitação, o que é um fator crítico na avaliação do risco de enchentes. A visualização também permite identificar a dispersão dos dados e possíveis comportamentos sazonais ou padrões extremos, fundamentais para alimentar modelos preditivos e sistemas de monitoramento inteligente.

3. ANÁLISE ESTATÍSTICA DESCRITIVA

Medida	Valor
Média	150,02
Mediana	150,62
Moda	219,00
Mínimo	0,01

Máximo	299,97
Amplitude	299,96
Variância	7401,53
Desvio Padrão	86,03
Coeficiente de Variação (%)	57,35
1º Quartil (Q1)	76,12
2º Quartil (Mediana)	150,62
3º Quartil (Q3)	223,40

(Tabela com as medidas estatísticas descritivas da variável Rainfall (mm))

A análise estatística da variável Rainfall (mm) revela uma média de 150,02 mm e uma mediana de 150,62 mm, indicando uma distribuição bastante simétrica, onde os valores centrais estão muito próximos. A moda de 219 mm sugere que esse é o valor de precipitação que ocorre com maior frequência nas localidades analisadas. A amplitude total de 299,96 mm, obtida pela diferença entre o valor mínimo (0,01 mm) e o máximo (299,97 mm), evidencia uma vasta variação no volume de chuva entre os registros.

Esse comportamento é reforçado pela variância de 7401,53 e pelo desvio padrão de 86,03, que demonstram uma dispersão considerável dos dados em torno da média. O coeficiente de variação de 57,35% classifica a variabilidade como alta, o que é característico de fenômenos climáticos, que tendem a apresentar flutuações significativas.

Além disso, a análise dos quartis mostra que 25% dos registros apresentam volume de chuva inferior a 76,12 mm (Q1), enquanto 50% dos dados estão abaixo de 150,62 mm (mediana) e 75% estão abaixo de 223,40 mm (Q3). Esses dados são fundamentais para compreender o comportamento da precipitação na região analisada e servem de base para a construção de modelos preditivos e sistemas de alerta voltados à prevenção de enchentes urbanas.

4. CONCLUSÃO

A análise estatística realizada sobre a base de dados de risco de enchentes permite compreender, de forma estruturada, o comportamento da variável precipitação (Rainfall (mm)) e sua relação direta com a ocorrência de enchentes nas localidades analisadas. A distribuição equilibrada entre registros com e sem enchentes, somada à expressiva variabilidade dos volumes de chuva, evidencia a complexidade dos fatores que impactam esses eventos.

Diante desse cenário, torna-se evidente a importância de desenvolver soluções tecnológicas inteligentes, capazes de processar grandes volumes de dados ambientais, identificar padrões e gerar previsões precisas. A utilização de análises estatísticas como etapa preliminar é fundamental para embasar a construção de modelos preditivos, alimentados tanto por dados históricos quanto por sensores em tempo real.

Portanto, esta análise não apenas fornece uma visão aprofundada dos dados climáticos, mas também serve como fundamento técnico e científico para o desenvolvimento de sistemas cognitivos aplicados à gestão de riscos urbanos, especialmente no monitoramento e na resposta a eventos extremos, como enchentes. A adoção de tecnologias baseadas em inteligência artificial, aliada à estatística descritiva e preditiva, é um caminho essencial para mitigar os impactos socioeconômicos causados pelas mudanças climáticas e proteger as comunidades urbanas.