



Construindo um futuro onde a IA  
melhora e avança a sociedade



# IA Aplicada aos Desafios Socioambientais da Amazônia

Onédio Junior



# Encontro 6

## *Vibe Coding para Análise de Dados*



## OBJETIVO DO 6º ENCONTRO

---

- Capacitar os alunos a utilizar a abordagem de *Vibe Coding* para conduzir um ciclo completo de análise de dados, desde a manipulação e visualização até a análise estatística inicial, aplicando esses conhecimentos em um desafio prático com dados socioambientais da Amazônia.



# MAPA DA TRILHA DO CONHECIMENTO

---

- Análise de Dados.
- *Vibe Coding* para o Ciclo de Análise de Dados.
- Desafio da Semana



# MAPA DA TRILHA DO CONHECIMENTO

---

- **Análise de Dados.**
- *Vibe Coding para o Ciclo de Análise de Dados.*
- Desafio da Semana.

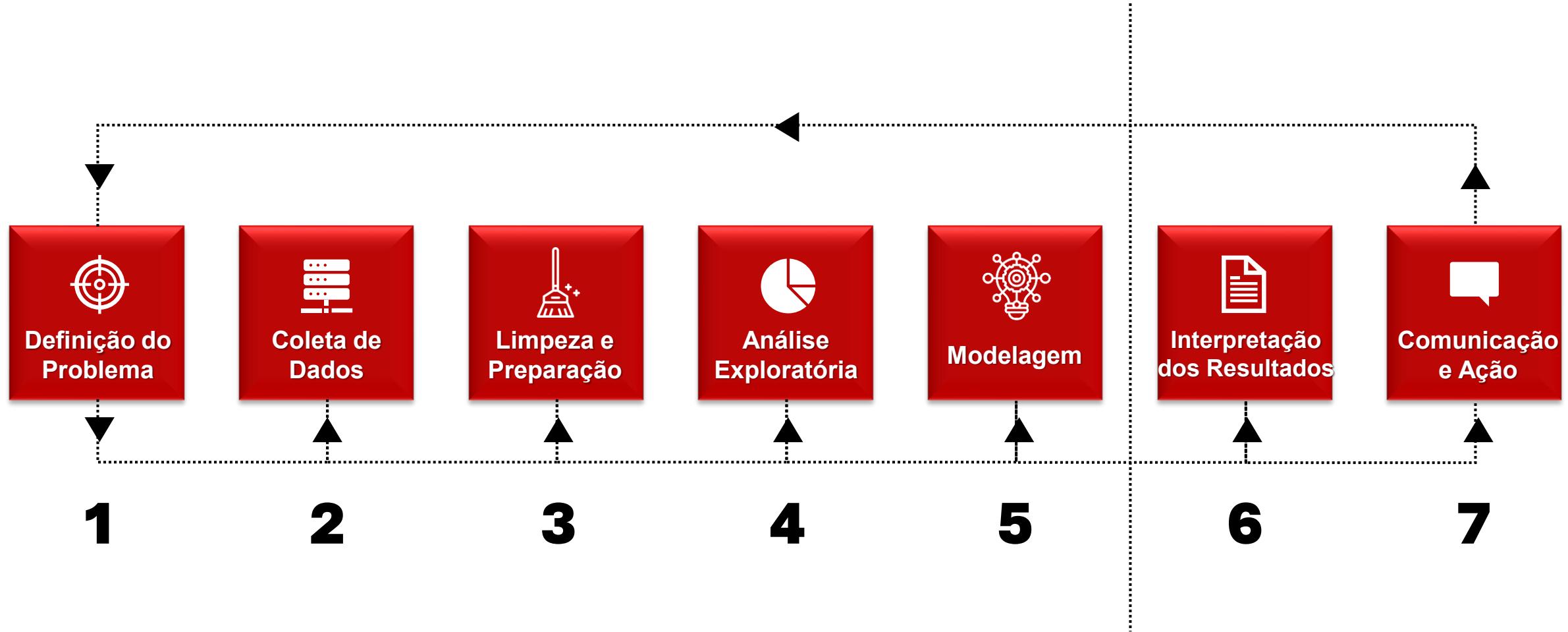
# COMO APRENDER IA

---



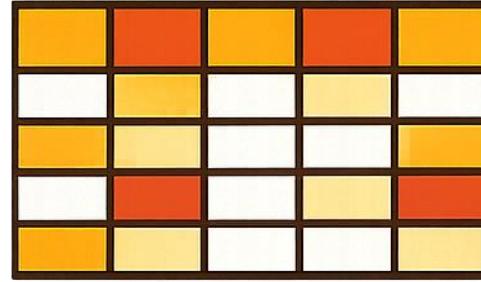


Análise de dados é o processo de **inspecionar, limpar, transformar e modelar dados** para **descobrir informações úteis, tirar conclusões** e apoiar a tomada de decisões.



**Insights valiosos extraídos de informações brutas.**

## DADOS ESTRUTURADOS



## DADOS NÃO ESTRUTURADOS

text media file  
messages  
blog video doc  
ideas email file  
unstructured



# LIMPEZA E ORGANIZAÇÃO DOS DADOS

id	Nome	Altura
1	Maria	57
2	Clara	59
3B	Roberta	58
6	Clara	59



id	Nome	Peso
1	Maria	57
2	Clara	59
3	Roberta	58



# EXPLORAÇÃO: DEIXEM OS DADOS FALAREM





# MAPA DA TRILHA DO CONHECIMENTO

---

- Análise de Dados.
- *Vibe Coding* para o Ciclo de Análise de Dados.
- Desafio da Semana.

## DEFINIÇÃO DO PROBLEMA/OBJETIVO

- Compreenda o problema de negócio com clareza.
- Formule perguntas específicas e mensuráveis.
- Defina métricas ou indicadores de sucesso.
- Considere o contexto e os stakeholders envolvidos.

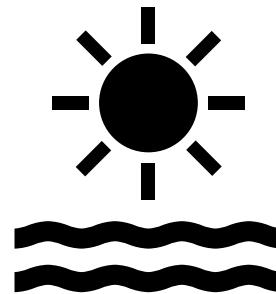


## CENÁRIO

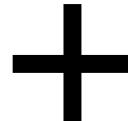
Em uma comunidade amazônica, agricultores familiares relatam perdas frequentes nas colheitas devido à falta de previsibilidade sobre o regime de chuvas. Apesar de utilizarem calendários tradicionais, os padrões climáticos mudaram nos últimos anos, afetando a produtividade e a segurança alimentar local. Os líderes da comunidade querem entender por que essas perdas estão aumentando e se é possível prever melhor os períodos críticos. Eles desejam identificar o problema real com clareza, buscando caminhos que ajudem a tomar decisões baseadas em evidências para garantir a sustentabilidade da produção.



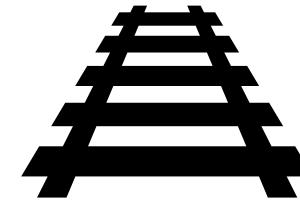
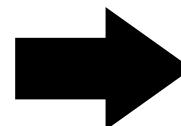
## VIBE CODING PARA DEFINIÇÃO DO PROBLEMA



Cenário



Informações  
Coletadas



Definir o  
Problema



## VIBE CODING PARA DEFINIÇÃO DO PROBLEMA

Você é um analista de dados que trabalha com uma equipe multidisciplinar em um projeto voltado à sustentabilidade e à segurança alimentar de comunidades amazônicas. Seu papel é ajudar a entender melhor as causas das perdas nas colheitas, agravadas por mudanças no regime de chuvas. Ajude-me a compreender com mais clareza qual é o problema central enfrentado por essa comunidade. Para isso, analise o cenário apresentado, identifique os desafios e proponha perguntas adicionais que aprofundem o entendimento da situação. "Em uma comunidade amazônica, agricultores familiares relatam perdas frequentes nas colheitas devido à falta de previsibilidade sobre o regime de chuvas. Apesar de utilizarem calendários tradicionais, os padrões climáticos mudaram nos últimos anos, afetando a produtividade e a segurança alimentar local. Os líderes da comunidade querem entender por que essas perdas estão aumentando e se é possível prever melhor os períodos críticos. Eles desejam identificar o problema real com clareza, buscando caminhos que ajudem a tomar decisões baseadas em evidências para garantir a sustentabilidade da produção.". Retorne um parágrafo curto explicando o possível problema central e, em seguida, liste 3 a 4 perguntas que um analista poderia fazer à comunidade para investigar melhor o problema.



## VIBE CODING PARA DEFINIÇÃO DO PROBLEMA

### Problema central:

O problema central enfrentado pela comunidade é a imprevisibilidade do regime de chuvas, que compromete a eficácia dos calendários agrícolas tradicionais e, com isso, impacta negativamente a produtividade e a segurança alimentar. Essa mudança nos padrões climáticos dificulta o planejamento das safras, levando a perdas recorrentes e insegurança quanto à produção sustentável.

### Perguntas para aprofundar o entendimento:

- 1) Quais sinais climáticos a comunidade costumava observar para iniciar o plantio?
- 2) Há registros locais ou percepções sobre mudanças no início ou intensidade das chuvas?

....

Escreva seu prompt

+  Ferramentas



## VIBE CODING PARA DEFINIR MÉTRICAS

Você é um analista de dados que trabalha com uma equipe multidisciplinar em um projeto voltado à sustentabilidade e à segurança alimentar de comunidades amazônicas. Seu papel é ajudar a entender melhor as causas das perdas nas colheitas, agravadas por mudanças no regime de chuvas. Ajude a identificar métricas que permitam avaliar se as ações propostas estão melhorando a previsibilidade e a produção agrícola. A comunidade sofre perdas nas colheitas e deseja usar dados para tomar decisões com base em evidências. Liste de 3 a 4 métricas claras, diretamente relacionadas ao problema identificado.



## COLETA DE INFORMAÇÕES



**Exemplo de perguntas a serem realizadas**

- 1) Quais meses costumavam ter mais chuvas e isso mudou recentemente?
- 2) Que sinais a comunidade observava antes das chuvas e ainda percebe hoje?
- 3) Como as perdas nas colheitas impactaram a alimentação e a renda da comunidade nos últimos anos?



## VIBE CODING PARA DEFINIR MÉTRICAS

Aqui estão quatro métricas que podem ser utilizadas para avaliar a situação da comunidade:

### 1. Desvio médio entre chuvas previstas e reais

→ Mede a precisão da previsão em relação à realidade.

### 2. Número de dias com chuvas fora do calendário tradicional

→ Ajuda a identificar mudanças no padrão climático local.

### 3. Percentual de perdas nas colheitas por estação

→ Avalia o impacto da imprevisibilidade climática na produção.

### 4. Rendimento médio da colheita (kg/hectare)

→ Permite verificar se ações baseadas em dados estão aumentando a produtividade.

Escreva seu prompt

+  Ferramentas



## COLETA DE DADOS

- Identifique as fontes confiáveis de dados internos e externos disponíveis para análise.
- Avalie a qualidade e a relevância dos dados antes de utilizá-los.
- Registre como e quando os dados foram coletados para garantir rastreabilidade.
- Diversifique os métodos de coleta: sensores, entrevistas, formulários, APIs, etc.
- Garanta o consentimento e a ética no uso de dados sensíveis da comunidade.



## COLETA DE DADOS



### Planeje o que será coletado

Defina as variáveis, fontes e formatos dos dados.



### Escolha os métodos de coleta

Use entrevistas, sensores, formulários, APIs ou bancos de dados.



### Organize o processo de registro

Documente local, data, responsáveis e instrumentos utilizados.



### Garanta segurança e consentimento

Proteja dados sensíveis e obtenha autorização da comunidade.

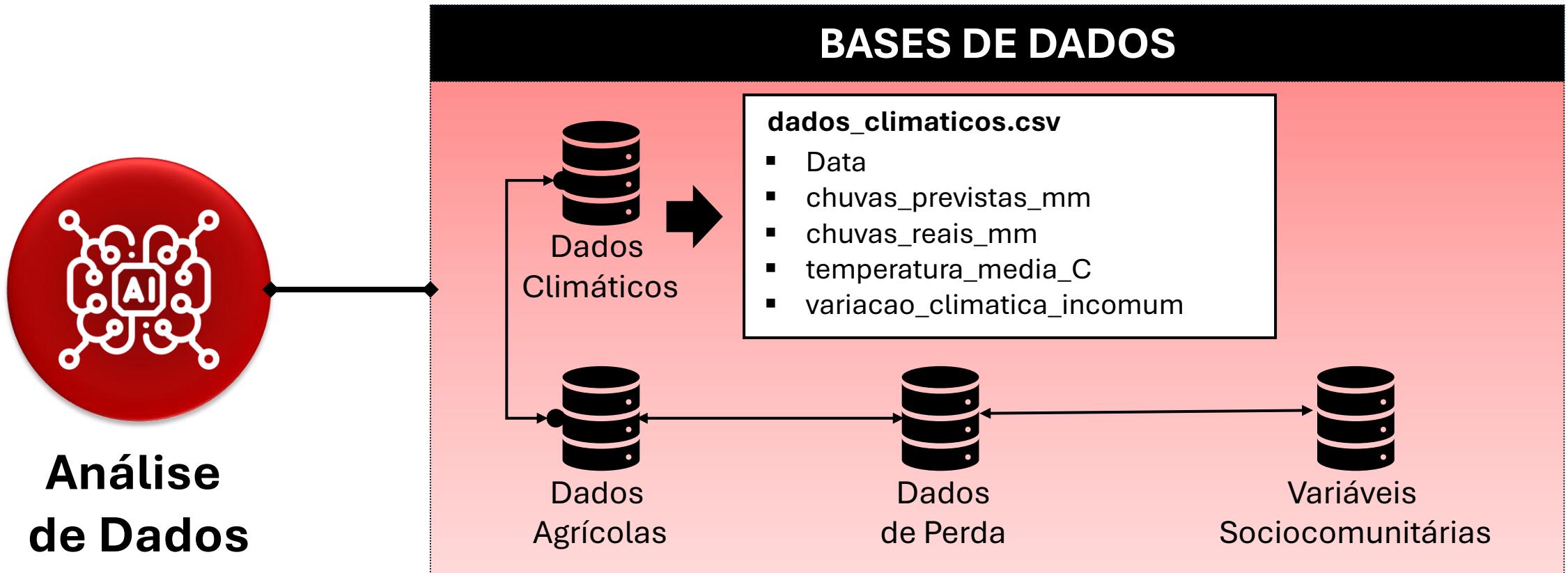


## LIMPEZA E PREPARAÇÃO

- Remova duplicatas para evitar distorções na análise.
- Trate valores ausentes com critérios claros e consistentes.
- Corrija erros de digitação e formatação nos dados.
- Padronize nomes, datas e categorias para uniformidade.
- Identifique e trate outliers com base no contexto.
- Converta dados para formatos adequados à análise futura.



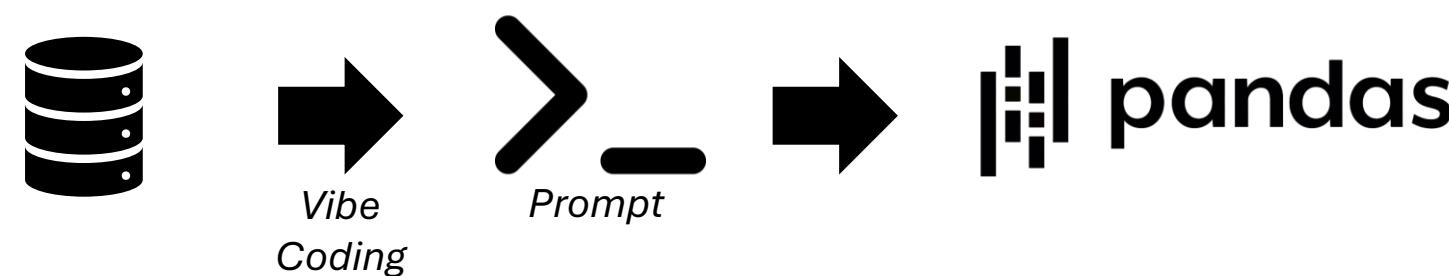
## CARREGANDO O DATASET



## VIBE CODING CARREGANDO O DATASET

```
data,chuvas_previstas_mm,chuvas_reais_mm,temperatura_media_C,variacao_climatica_incomum
2024-02-20,12,11.5,25.1,nao
2024-02-21,15,14.7,24.9,não
2024-02-22,13,13.2,,não
2024-02-23,17,17.3,25.0,sim
2024-02-24,12,12.1,25.2,sim
2024-02-25,9,8.6,25.0,sim
2024-02-26,13,13.0,24.8,sim
2024-02-27,10,10.1,24.9,não
2024-02-28,16,16.1,25.1,não
...

```



## VIBE CODING CARREGANDO O DATASET

Você é um analista de dados com experiência em análise ambiental e sustentabilidade, atuando em projetos voltados à Amazônia. Precisa iniciar a exploração de um dataset com registros climáticos que servirão de base para investigar padrões de chuvas e variações climáticas locais. Carregue o arquivo CSV em um DataFrame do Pandas, chamado df\_dados\_climaticos, garantindo que datas sejam interpretadas corretamente. O arquivo anexo chama-se dados\_climaticos.csv e contém colunas como **Data**, **Chuvas\_previstas\_mm**, **Chuvas\_reais\_mm**, **Temperatura\_media\_C**, e **Variação\_climática\_incomum**. Apresente as 5 primeiras linhas do DataFrame na tela.



## VIBE CODING CARREGANDO O DATASET

data	chuvas_previstas_mm	chuvas_reais_mm	temperatura_media_C	variacao_climatica_incomum
2024-02-20	12.0	11.5	25.1	nao
2024-02-21	15.0	14.7	24.9	não
2024-02-22	13.0	13.2	NaN	não
2024-02-23	17.0	17.3	25.0	sim
2024-02-24	12.0	12.1	25.2	sim



## VIBE CODING PARA VERIFICAR DUPLICATAS

Você é um analista de dados atuando em um projeto de sustentabilidade e segurança alimentar na Amazônia. Seu objetivo é garantir que o conjunto de dados climáticos reflita com precisão a realidade local, evitando distorções provocadas por registros duplicados. Analise o DataFrame df\_dados\_climaticos e verifique se há registros duplicados. Identifique quantas duplicatas existem e, se possível, exiba os primeiros registros duplicados encontrados. O DataFrame já foi carregado e se chama df\_dados\_climaticos. A saída deverá atender ao exemplo abaixo:

Número de registros duplicado : XX registros



## VIBE CODING VALORES AUSENTES



*O que você faria com 10% de dados faltando na estação de monitoramento de chuvas?*



## ESTRATÉGIAS PARA VALORES AUSENTES



### Remoção de registros

Exclui linhas com valores nulos quando impacto for pequeno.



### Preenchimento com média

Substitui valor ausente pela média da coluna numérica.



### Preenchimento com mediana

Usa mediana para evitar distorção por outliers nos dados.



### Interpolação linear

Estima valores ausentes com base na sequência temporal dos dados.



## VIBE CODING PARA VALORES AUSENTES

Você é um analista de dados atuando em um projeto de sustentabilidade e segurança alimentar na Amazônia. Seu objetivo é garantir que o conjunto de dados climáticos reflita com precisão a realidade local, evitando distorções provocadas por ausência de dados. Analise o DataFrame df\_dados\_climaticos e verifique se há ausência de dados. **Identifique quais variáveis apresentam ausência de valores**, e traga o resultado apresentando uma tabela mostrando o nome da coluna e o número de valores ausentes.



## VIBE CODING PARA VALORES AUSENTES

Valores Ausentes No Dataset



	Coluna	Valores_Ausentes
1	chuvas_previstas_mm	4
2	chuvas_reais_mm	8
3	temperatura_media_C	1



## VIBE CODING PARA VALORES AUSENTES

Substitua os valores ausentes nas colunas ***chuvas\_previstas\_mm***, ***chuvas\_reais\_mm*** e ***temperatura\_media\_C*** do DataFrame `df_dados_climaticos` utilizando a **média de cada coluna**. Certifique-se de aplicar a substituição apenas se houver valores ausentes.

Informe, para cada coluna, se havia valores ausentes e se eles foram substituídos.

Exemplo:  
`chuvas_previstas_mm`: 5 valores ausentes substituídos.

`chuvas_reais_mm`: Nenhuma modificação necessária.

`temperatura_media_C`: 2 valores ausentes substituídos.



## VIBE CODING PARA VALORES AUSENTES

**Liste todos os valores únicos existentes** na coluna *variacao\_climatica\_incomum* do DataFrame df\_dados\_climaticos, acompanhados da respectiva quantidade de ocorrências de cada valor.

Formato da saída:

sim : XX valores

não : XX valores

nao : XX valores

(outros valores, se houver, seguindo o mesmo padrão)



## VIBE CODING PARA VALORES AUSENTES

Padronize os valores da coluna *variacao\_climatica\_incomum* do DataFrame df\_dados\_climaticos, substituindo todas as variações como "nao", "Nao", "não", "NAO" por "não", garantindo que a coluna contenha apenas os valores "sim" ou "não", em minúsculas e com acento, conforme o português correto.

Retorne os valores únicos resultantes da coluna após a padronização, seguidos da quantidade de ocorrências, conforme o formato abaixo:

sim : XX valores

não : XX valores



## VIBE CODING PARA VALORES AUSENTES

**sim** : 51 valores

**não** : 100 valores

Todos os valores da coluna ***variacao\_climatica\_incomum*** foram padronizados corretamente para apenas "sim" ou "não".



## VIBE CODING PARA VALORES AUSENTES

**sim** : 51 valores

**não** : 100 valores

Todos os valores da coluna ***variacao\_climatica\_incomum*** foram padronizados corretamente para apenas "sim" ou "não".



## VIBE CODING PARA EXPORTAR O BD LIMPO

Exporte o DataFrame `df_dados_climaticos` já tratado para um arquivo CSV chamado `dados_climaticos_tratado.csv`. Em seguida, disponibilize o link para download direto do arquivo. Forneça apenas uma linha com o link completo para download do arquivo CSV gerado, sem comentários ou explicações adicionais.



## ANÁLISE EXPLORATÓRIA

- Exploração inicial dos dados com gráficos e estatísticas.
- Identificação de padrões e tendências relevantes.
- Avaliação de distribuições e variáveis categóricas.
- Detecção de outliers e inconsistências visuais.
- Criação de hipóteses para testar futuramente.
- Suporte à próxima fase: modelagem e decisões



## VIBE CODING PARA ESTRUTURA DOS DADOS

**Verifique a estrutura do DataFrame df\_dados\_climaticos, informando a quantidade de variáveis (colunas), o número de instâncias (linhas) e o tipo de dado associado a cada coluna.**

**Formato da saída:**

Total de linhas: [número]

Total de colunas: [número]

Tipos de dados por coluna:

coluna\_1: tipo

coluna\_2: tipo

coluna\_n: tipo



## VIBE CODING PARA ESTRUTURA DOS DADOS

Total de linhas: 151

Total de colunas: 5

### Tipos de dados por coluna:

data: datetime64[ns]

chuvas\_previstas\_mm: float64

chuvas\_reais\_mm: float64

temperatura\_media\_C: float64

variacao\_climatica\_incomum: object



## VIBE CODING PARA ESTRUTURA DOS DADOS

**Verifique a estrutura do DataFrame df\_dados\_climaticos, informando a quantidade de variáveis (colunas), o número de instâncias (linhas) e o tipo de dado associado a cada coluna.**

**Formato da saída:**

Total de linhas: [número]

Total de colunas: [número]

Tipos de dados por coluna:

coluna\_1: tipo

coluna\_2: tipo

coluna\_n: tipo



## VIBE CODING PARA ESTATÍSTICA DESCRIPTIVA

Apresente as estatísticas descritivas das variáveis numéricas do DataFrame df\_dados\_climaticos, incluindo média, mediana, mínimo, máximo e desvio padrão para cada coluna.

Formato da saída:

Coluna: <nome\_da\_coluna>

Média: <valor>

Mediana: <valor>

Mínimo: <valor>

Máximo: <valor>

Desvio padrão: <valor>

---

(repetir para cada coluna numérica)



## VIBE CODING PARA ESTATÍSTICA DESCRIPTIVA

Coluna: chuvas\_previstas\_mm

Média: 16.09

Mediana: 14.00

Mínimo: 9.00

Máximo: 26.70

Desvio padrão: 5.09

...







Construindo um futuro onde a IA  
melhora e avança a sociedade