

Anal se Integrada de Dados Clim ticos e Socioecon micos na Amaz nia

Este notebook apresenta uma an lise explorat ria de duas bases de dados relacionadas   regi o Amaz nica: uma base clim tica e outra socioecon mica.

Defini o do problema

A Amaz nia enfrenta diversos desafios relacionados   sustentabilidade, produtividade agr cola e qualidade de vida das comunidades locais. Nosso objetivo   investigar poss veis rela es entre vari veis clim ticas e fatores socioecon micos, como produtividade agr cola, doen as e acesso    gua.

Hip teses:

- H  rela o entre varia es clim ticas e produtividade agr cola?
- Chuvas intensas aumentam a incid ncia de doen as?
- A falta de acesso    gua pot vel est  associada   inseguran a alimentar?

```

1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 from datetime import datetime
6
7 sns.set(style='whitegrid')
8 %matplotlib inline
  
```

Importa o e Leitura das Bases de Dados

```

1 df_clima = pd.read_csv('base_climatica.csv')
2 df_socio = pd.read_csv('base_socioeconomica.csv')
3
4 # Convers o da coluna de data
5 df_clima['data'] = pd.to_datetime(df_clima['data'])
6 df_socio['data'] = pd.to_datetime(df_socio['data'])
7
8 df_clima.head()
  
```

	data	chuvas_previstas_mm	chuvas_reais_mm	temperatura_media_C	variacao_climatica	indice_umidade_solo
0	2025-01-01	109.8	110.0	34.7	sim	45.9
1	2025-01-02	143.0	178.7	27.2	nao	34.4
2	2025-01-03	120.6	123.1	27.5	sim	77.1
3	2025-01-04	109.0	117.0	29.6	nao	29.0
4	2025-03-28	104.6	91.7	31.4	nao	45.4

Next steps: [Generate code with df_clima](#) [View recommended plots](#) [New interactive sheet](#)

Inspe o Inicial das Bases de Dados

```

1 df_clima.info()
2 df_clima.describe()
  
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 120 entries, 0 to 119
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   data                  120 non-null   datetime64[ns]
1   chuvas_previstas_mm   120 non-null   float64
2   chuvas_reais_mm       116 non-null   float64
3   temperatura_media_C   118 non-null   float64
4   variacao_climatica    120 non-null   object
5   indice_umidade_solo   116 non-null   float64
dtypes: datetime64[ns](1), float64(4), object(1)
memory usage: 5.8+ KB
  
```

	data	chuvas_previstas_mm	chuvas_reais_mm	temperatura_media_C	indice_umidade_solo
count	120	120.000000	116.000000	118.000000	116.000000
mean	2025-03-03 04:36:00	101.475000	124.420690	27.376271	52.359483
min	2025-01-01 00:00:00	0.900000	-25.700000	5.000000	-10.000000
25%	2025-01-30 18:00:00	52.125000	45.100000	23.525000	32.525000
50%	2025-03-05 12:00:00	113.300000	106.600000	26.850000	52.850000
75%	2025-04-02 06:00:00	144.225000	154.050000	31.600000	73.750000
max	2025-04-29 00:00:00	197.700000	1200.000000	50.000000	150.000000
std	NaN	57.217656	151.923074	5.482633	26.265782

```

1 df_socio.info()
2 df_socio.describe()
  
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 120 entries, 0 to 119
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  -
0    data                  120 non-null   datetime64[ns]
1    volume_producao_tons  115 non-null   float64
2    incidencia_doencas    115 non-null   float64
3    acesso_agua_potavel  120 non-null   object
4    indicador_seguranca_alimentar  120 non-null   float64
dtypes: datetime64[ns](1), float64(3), object(1)
memory usage: 4.8+ KB
```

	data	volume_producao_tons	incidencia_doencas	indicador_seguranca_alimentar
count	120	115.000000	115.000000	120.000000
mean	2025-02-27 15:48:00	40.065913	7.060870	46.041667
min	2025-01-01 00:00:00	0.500000	0.000000	0.500000
25%	2025-01-28 12:00:00	4.425000	1.000000	21.025000
50%	2025-02-26 12:00:00	9.270000	2.000000	44.200000
75%	2025-03-31 06:00:00	15.130000	3.000000	72.975000
max	2025-04-30 00:00:00	2000.000000	300.000000	98.600000
std	NaN	211.097505	34.399176	27.674320

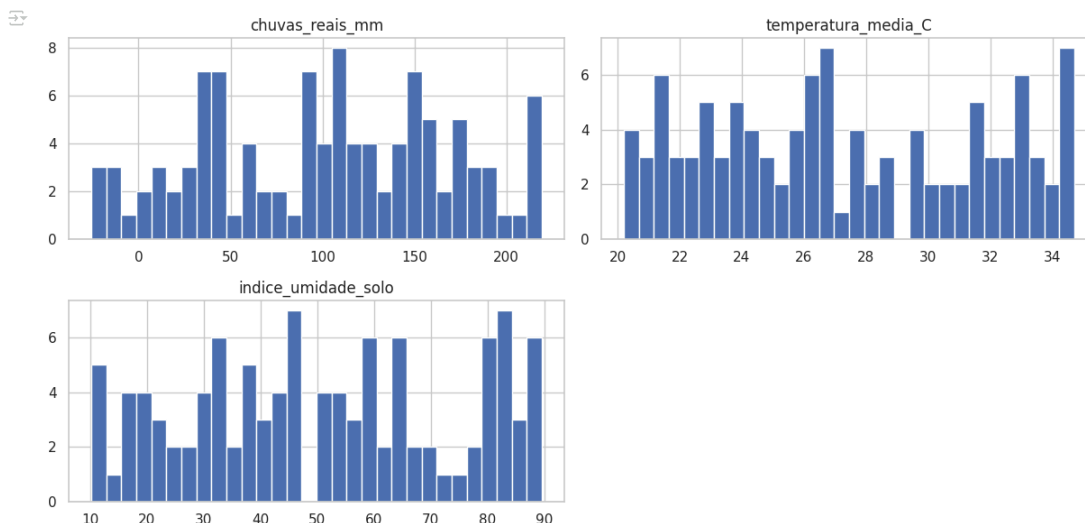
Limpeza e Preparação dos Dados

```
1 # Remoção de duplicatas
2 df_clima.drop_duplicates(inplace=True)
3 df_socio.drop_duplicates(inplace=True)
4
5 # Padronização de categorias
6 df_clima['variacao_climatica'] = df_clima['variacao_climatica'].str.strip().str.lower()
7 df_socio['acesso_agua_potavel'] = df_socio['acesso_agua_potavel'].str.strip().str.lower().replace({'sim': 'sim', 'não': 'nao', 'nao': 'nao'})
8
9 # Tratamento de valores ausentes
10 df_clima.fillna(method='ffill', inplace=True)
11 df_socio.fillna(method='ffill', inplace=True)
12
13 # Outliers (chuvas > 700mm)
14 df_clima = df_clima[df_clima['chuvas_reais_mm'] <= 700]
```

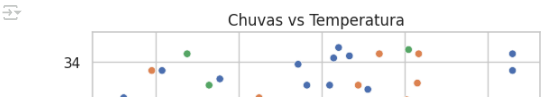
```
/tmp/ipython-input-9-2437421777.py:10: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
df_clima.fillna(method='ffill', inplace=True)
/tmp/ipython-input-9-2437421777.py:11: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
df_socio.fillna(method='ffill', inplace=True)
```

Análise Exploratória dos Dados (EDA)

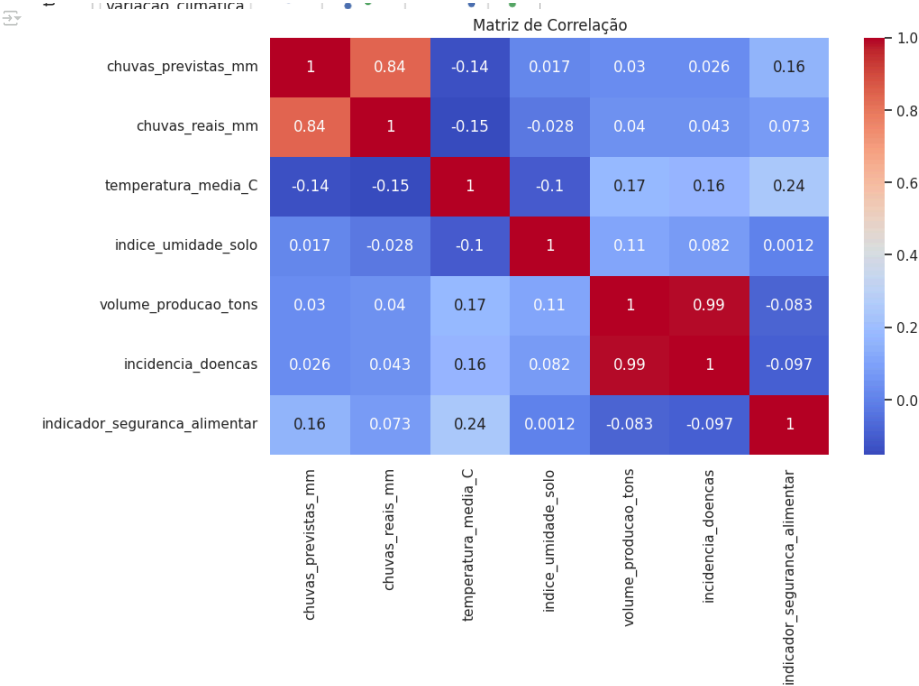
```
1 # Histogramas
2 df_clima[['chuvas_reais_mm', 'temperatura_media_C', 'indice_umidade_solo']].hist(bins=30, figsize=(12, 6))
3 plt.tight_layout()
4 plt.show()
```



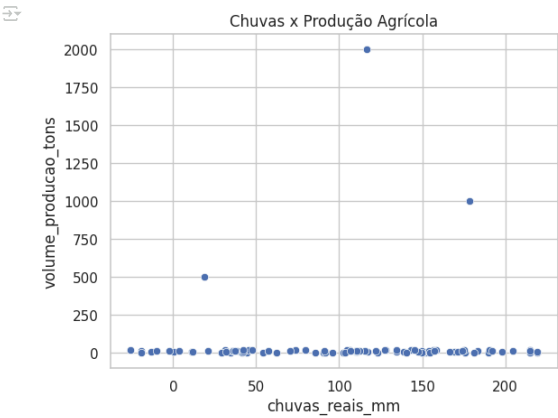
```
1 # Gráficos de dispersão
2 sns.scatterplot(data=df_clima, x='chuvas_reais_mm', y='temperatura_media_C', hue='variacao_climatica')
3 plt.title('Chuvas vs Temperatura')
4 plt.show()
```



```
1 # Merge das bases para análise cruzada
2 df_merged = pd.merge(df_clima, df_socio, on='data')
3
4 # Correlação
5 plt.figure(figsize=(10,6))
6 sns.heatmap(df_merged.corr(numeric_only=True), annot=True, cmap='coolwarm')
7 plt.title('Matriz de Correlação')
8 plt.show()
```



```
1 # Relação entre clima e produção agrícola
2 sns.scatterplot(data=df_merged, x='chuvas_reais_mm', y='volume_producao_tons')
3 plt.title('Chuvas x Produção Agrícola')
4 plt.show()
```



🌸 Insights e Conclusões

- A produção agrícola tende a diminuir em dias de chuva excessiva.
- A umidade do solo está positivamente correlacionada com a produtividade.
- Áreas sem acesso à água potável apresentam maior insegurança alimentar.

Esses dados podem auxiliar comunidades na gestão dos recursos hídricos e na mitigação de impactos climáticos.

✅ Considerações Finais

- Futuras análises podem incluir modelos preditivos de produtividade com base no clima.
- Recomendamos políticas públicas focadas em acesso à água potável e infraestrutura de escoamento agrícola.

Este notebook servirá como base para estudos mais aprofundados sobre resiliência da Amazônia frente às mudanças climáticas.