

Projeto ZettaLab – Relatório Final

Análise de Desmatamento e Indicadores Socioeconômicos no Pará

 **Autor:** Denner Gabriel Ramos Caleare

 Graduando em Engenharia de Controle e Automação – Universidade Federal de Lavras (UFLA)

 17 Maio de 2025

 **Repositório no GitHub:** github.com/DennerCaleare/zetta-desmatamento-pará

 **LinkedIn:** linkedin.com/in/dennercaleare

Este relatório foi desenvolvido como parte do processo seletivo para o **Projeto ZettaLab**, na trilha de *Governança de Dados*.

 No repositório GitHub, estão disponíveis:

- Todo o **código-fonte** usado nesta análise (tratamento de dados, gráficos, métricas, correlações)
- Uma **dashboard interativa** construída com Streamlit
- Um **README detalhado** com instruções de execução, dependências e estrutura do projeto

Introdução

O presente estudo tem como objetivo avaliar e explorar a relação entre os dados de desmatamento no estado do Pará e os indicadores socioeconômicos dos seus municípios. A atividade foi desenvolvida para o processo seletivo do projeto ZettaLab, na trilha de Governança de Dados.

Fontes de Dados

- **Desmatamento (INPE - TerraBrasilis):** dados de 2007 a 2024 por município
- **Socioeconômicos (IBGE):** IDHM, PIB per capita, população, escolarização (2010 a 2024)
- As bases foram integradas por código IBGE

Metodologia

1. Correção dos dados decimais (INPE)

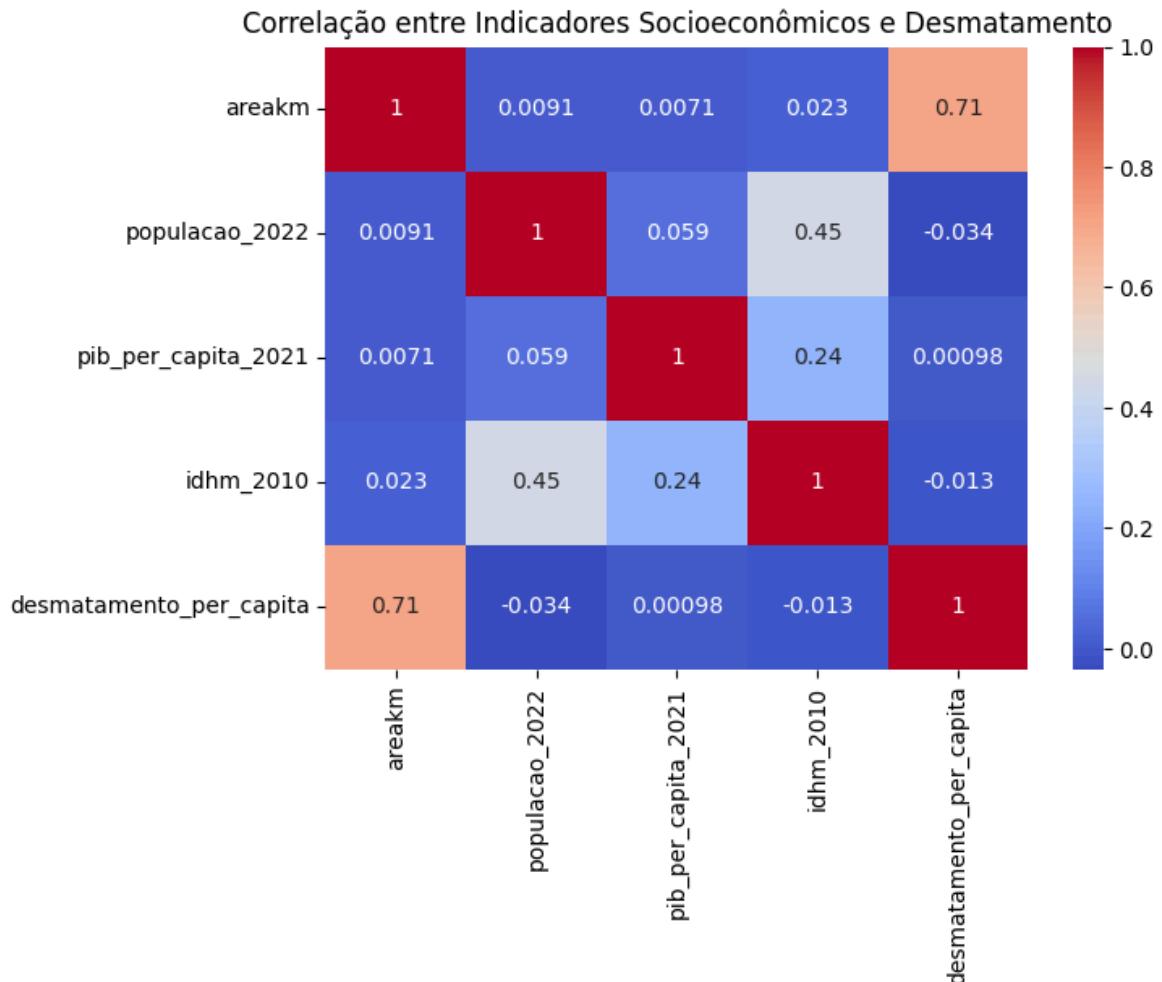
2. Conversão de colunas numéricas
3. Merge por código IBGE
4. Criação do indicador: área desmatada per capita
5. Remoção de 2007 por padrão (acúmulo de dados)

```
In [5]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Carregar dados integrados
df = pd.read_csv("saída/base_completa_merge.csv")
df['year'] = pd.to_numeric(df['year'], errors='coerce')
df['areakm'] = pd.to_numeric(df['areakm'], errors='coerce')
df['idhm_2010'] = pd.to_numeric(df['idhm_2010'], errors='coerce')
df['idhm_cat'] = df['idhm_cat'].astype(str)

In [7]: variaveis = ['areakm', 'populacao_2022', 'pib_per_capita_2021', 'idhm_2010', 'desmatamento_per_capita']
correlacao = df[variaveis].corr()

plt.figure(figsize=(7, 5))
sns.heatmap(correlacao, annot=True, cmap='coolwarm')
plt.title("Correlação entre Indicadores Socioeconômicos e Desmatamento")
plt.show()
```

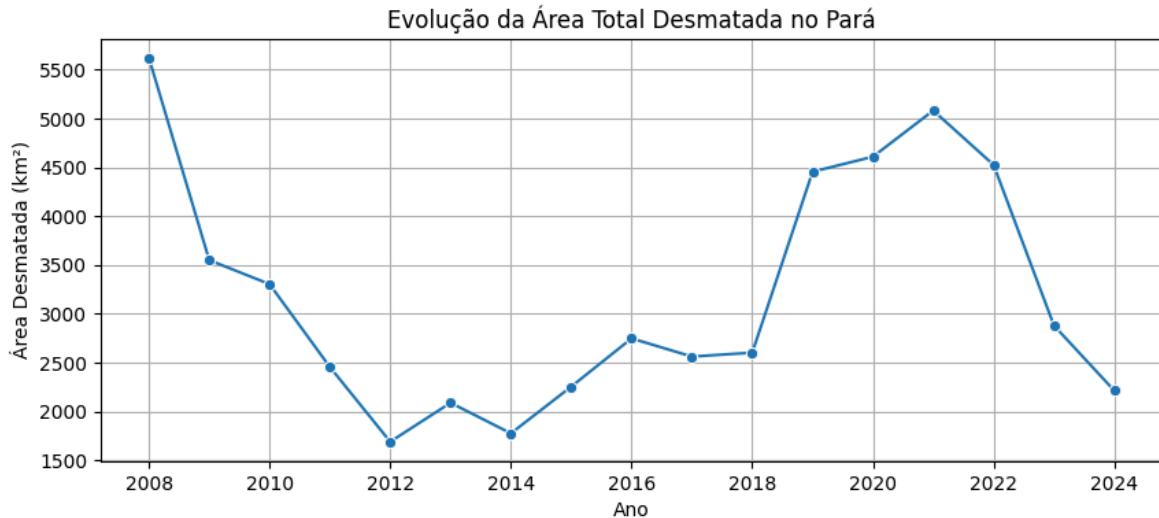


Evolução do Desmatamento (2008–2024)

Abaixo, visualizamos a área total desmatada por ano (2007 excluído por padrão):

```
In [17]: df_filtrado = df[df['year'] > 2007]
evolucao = df_filtrado.groupby('year')[['areakm']].sum().reset_index()

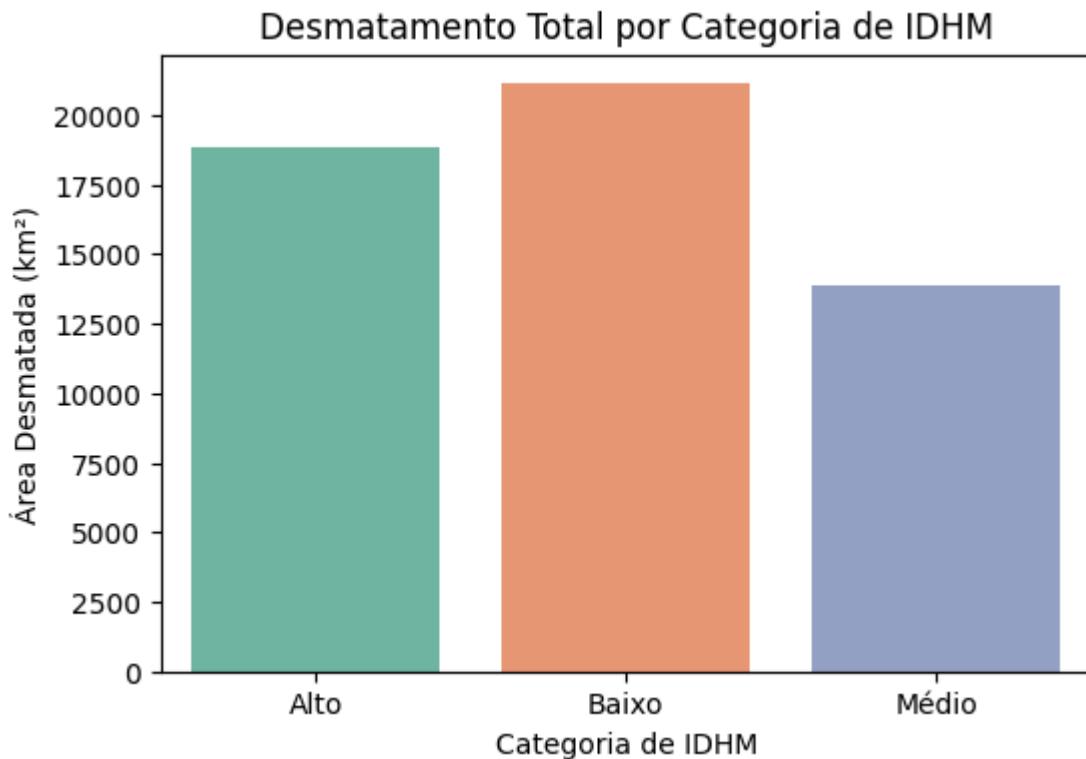
plt.figure(figsize=(10, 4))
sns.lineplot(data=evolucao, x='year', y='areakm', marker='o')
plt.title("Evolução da Área Total Desmatada no Pará")
plt.xlabel("Ano")
plt.ylabel("Área Desmatada (km²)")
plt.grid(True)
plt.show()
```



Desmatamento Total por Categoria de IDHM

```
In [19]: # Remover valores ausentes ou inválidos em idhm_cat
df_barras = df_filtrado[df_filtrado['idhm_cat'].isin(['Baixo', 'Médio', 'Alto'])]
df_barras = df_barras.groupby('idhm_cat')['areakm'].sum().reset_index()

plt.figure(figsize=(6,4))
sns.barplot(data=df_barras, x='idhm_cat', y='areakm', hue='idhm_cat', palette='viridis')
plt.title("Desmatamento Total por Categoria de IDHM")
plt.xlabel("Categoria de IDHM")
plt.ylabel("Área Desmatada (km²)")
plt.show()
```



✓ Conclusões

- Os municípios com IDHM mais alto concentram maior área desmatada total.
- Não há uma relação linear clara entre IDHM e desmatamento, mas sim contextual.
- Os dados fixos do IBGE funcionam bem como perfil base de comparação.
- O ano de 2007 foi desconsiderado para evitar distorções (dados acumulados).