

Zetta Lab 2025, Desafio 01: Ciência e Governança de Dados

Participante: Fernando José Gregatti Noronha

Graduando em Ciência da Computação

Matrícula: 202120495

RELATÓRIO

1. INTRODUÇÃO

O relatório apresenta os resultados do Desafio 01 proposto pela Zetta Lab 2025. A atividade teve como objetivo principal explorar e analisar dados socioeconômicos e ambientais de municípios do Estado do Pará. Para a solução desta atividade, o escopo do problema foi reduzido aos anos de 2000 e 2010. A partir da análise desses dados, buscou-se compreender possíveis relações entre indicadores de desenvolvimento humano e os níveis de desmatamento registrados no território paraense.

A aplicação de técnicas de análise estatística foi essencial para identificar padrões, tendências e possíveis correlações entre os indicadores socioeconômicos e ambientais ao longo do tempo.

2. BASE DE DADOS

Para a realização deste estudo, foram selecionadas duas bases de dados públicas que oferecem informações relevantes sobre os municípios do Estado do Pará, abrangendo aspectos ambientais e socioeconômicos.

2.1 Dados de Desmatamento – INPE

A primeira base de dados utilizada refere-se às informações sobre desmatamento nos municípios paraenses. Esta base é mantida pelo Instituto Nacional de Pesquisas Espaciais (INPE) e está disponível na plataforma Base dos Dados. Ela fornece dados anuais sobre a extensão do desmatamento, permitindo análises temporais e espaciais.

As principais colunas presentes nesta base são:

- ano: ano de referência dos dados;
- id_municipio: código identificador do município;
- id_municipio_nome: nome do município;
- bioma: bioma predominante no município;
- area_total: área total do município;
- desmatado: área desmatada no ano de referência.
- vegetacao_natural: área de vegetação natural remanescente.

Esta base de dados é essencial para entender a dinâmica do desmatamento no Pará ao longo do tempo.

Link:

<https://basedosdados.org/dataset/1dd6317a-8e79-4a67-bb78-bc5fef214b4d?table=8dce525d-07f9-4d52-8ea0-36fa288d7df8>

2.2 Dados Socioeconômicos – IPEA

A segunda base de dados utilizada aborda indicadores socioeconômicos dos municípios do Pará. Esta base é fornecida pelo Instituto de Pesquisa Econômica Aplicada (IPEA) e também está disponível na plataforma Base dos Dados. Ela oferece uma visão abrangente sobre diversos aspectos sociais e econômicos dos municípios, permitindo análises comparativas e correlacionais. As principais colunas presentes nesta base são:

- ano: ano de referência dos dados;
- sigla_uf: sigla da unidade federativa;
- sigla_uf_nome: nome da unidade federativa;
- id_municipio: código identificador do município;
- id_municipio_nome: nome do município;
- raca_cor: classificação racial ou étnica;
- sexo: gênero;
- localizacao: localização urbana ou rural;
- ivs: Índice de Vulnerabilidade Social;
- idhm: Índice de Desenvolvimento Humano Municipal;
- prosperidade_social: indicador de prosperidade social;
- renda_per_capita: renda per capita do município;
- proporcao_vulneravel: proporção da população em situação de vulnerabilidade;
- taxa_envelhecimento: taxa de envelhecimento da população;
- populacao_total: população total do município.

Esta base de dados é fundamental para analisar o desenvolvimento humano e social dos municípios paraenses.

Link:

<https://basedosdados.org/dataset/1dd6317a-8e79-4a67-bb78-bc5fef214b4d?table=8dce525d-07f9-4d52-8ea0-36fa288d7df8>

3. LEITURA E TRATAMENTO DE DADOS

Após a obtenção das bases de dados, foi necessário realizar um processo de tratamento para garantir a consistência, integridade e usabilidade das informações.

Inicialmente, foi feita a leitura dos dados e, em seguida, a conversão dos tipos de variáveis para formatos apropriados. Em muitas bases públicas, os dados podem vir com tipos incorretos ou inconsistências causadas por leituras automatizadas. Para lidar com isso, os valores foram padronizados e convertidos para tipos compatíveis com as análises pretendidas.

Outro passo foi a remoção de colunas irrelevantes, ou seja, variáveis que não agregavam valor à análise e poderiam dificultar a visualização dos dados mais importantes. Paralelamente, os nomes das colunas foram renomeados de forma padronizada.

Além disso, foram realizadas verificações de duplicidade. Entradas duplicadas em bases de dados podem causar distorções significativas em análises estatísticas, especialmente quando envolvem contagem de ocorrências ou cálculos agregados.

Por fim, foram feitas verificações complementares para garantir a integridade da base unificada. Isso incluiu a checagem de valores ausentes, inconsistências ou erros remanescentes que pudessem comprometer a análise exploratória posterior.

4. ANÁLISE

Nesta etapa do estudo, foram realizadas análises estatísticas descritivas com o objetivo de entender melhor os dados socioeconômicos e ambientais dos municípios nos anos de 2000 e 2010. As variáveis analisadas incluíram área desmatada, Índice de Desenvolvimento Humano Municipal (IDHM), Índice de Vulnerabilidade Social (IVS), renda per capita e população. As análises incluíram o cálculo da média e desvio padrão dessas variáveis.

Os resultados mostraram um aumento no desmatamento médio, de 1333,65 km² em 2000 para 1778,64 km² em 2010, indicando uma expansão significativa da área desmatada, além de uma maior variação entre os municípios em 2010. No que se refere ao IDHM, observou-se uma melhoria considerável, com um aumento de 373,62 para 539,10, refletindo avanços nas condições de vida, principalmente em termos de educação, saúde e renda. Já o IVS apresentou uma queda, de 593,63 para 499,29, sugerindo um progresso na redução da vulnerabilidade social. Por outro lado, a renda per capita aumentou de 22.139,67 em 2000 para 28.652,50 em 2010, embora com uma maior desigualdade na distribuição dessa renda, como mostrado pelo aumento do desvio padrão.

Além disso, a população total dos municípios cresceu de 92.157,67 em 2000 para 117.318,15 em 2010, com grandes variações entre os municípios.

Gráficos comparativos também foram gerados para ilustrar essas variações ao longo do período de análise. Os gráficos mostram um aumento no desmatamento, e revelam avanços no IDHM, redução do IVS, crescimento da renda per capita e aumento da população.

Por fim, foi gerado um relatório de correlação para examinar as relações entre as variáveis socioeconômicas e ambientais. A análise revelou que, em 2000, algumas variáveis apresentaram correlações significativas, como entre a área total e a vegetação natural, e entre o IVS e a renda per capita. Em 2010, a correlação entre área total e desmatamento aumentou, o que pode indicar um vínculo mais forte entre o tamanho do município e o grau de desmatamento. A matriz de correlação gerada por meio do ProfileReport da biblioteca `pandas_profiling` permitiu uma visão detalhada das relações entre as variáveis, oferecendo insights valiosos para a compreensão das dinâmicas socioeconômicas e ambientais dos municípios ao longo do tempo.

5. CONCLUSÃO

Com base na análise realizada, pode-se concluir que, embora tenha havido avanços no Índice de Desenvolvimento Humano Municipal (IDHM) de diversos municípios do Pará, o crescimento do desmatamento no período de 2000 a 2010 levanta questões sérias sobre a sustentabilidade dos processos de desenvolvimento social e econômico na região. A correlação entre o aumento do IDHM e a intensificação do desmatamento sugere que o progresso em áreas como educação, saúde e renda pode estar sendo alcançado à custa de graves danos ambientais. Este cenário evidencia a urgência de se adotar políticas públicas que integrem práticas sustentáveis de uso da terra e desenvolvimento econômico, a fim de garantir que o crescimento social não comprometa irreversivelmente os recursos naturais da Amazônia.