

RELATÓRIO – ANÁLISE EXPLORATÓRIA DE DADOS

Disciplina: Ciências de Dados – Atividade 2

Autores: Adriano Sales Franco 418148

Edicleiton Alves da Silva 472590

João Marcos da Silva Leitão 377566

Luiz Matheus Sales Souza 586937

Ano: 2025

1. Objetivo Geral

O objetivo deste trabalho foi realizar uma análise exploratória de diferentes datasets disponibilizados no repositório para identificar padrões, tendências e relações entre variáveis socioeconômicas, demográficas e administrativas no Brasil. A partir das análises, buscou-se formular insights que auxiliem na criação de hipóteses, interpretações e potenciais aplicações práticas, tanto sociais quanto financeiras.

Foram utilizados quatro conjuntos de dados:

Atlas.xlsx – indicadores do IDHM por território.

PIMPF – Produção Física Industrial.xlsx – dados de produção industrial por segmentos.

ServidorPorOrgao.csv – distribuição de servidores federais, inativos e pensionistas.

2. Perguntas e Hipóteses Propostas

Para guiar o relatório e justificar os gráficos, foram definidas oito perguntas investigativas, cada uma com uma hipótese associada e um propósito social/econômico.

1. Como evoluiu o IDHM entre 1991 e 2010 nas unidades federativas?

Hipótese: Houve crescimento constante em todas as regiões, mas com intensidades distintas.

Propósito: Identificar desigualdades regionais e orientar políticas de investimento.

2. A dimensão “Renda” do IDHM acompanha proporcionalmente a evolução do IDHM geral?

Hipótese: A renda tem maior influência no crescimento geral do índice.

Propósito: Orientar estratégias de combate à pobreza e promoção de inclusão produtiva.

3. Quais estados apresentam maior discrepância entre Longevidade e Educação?

Hipótese: Alguns estados do Norte e Nordeste apresentam melhor longevidade que educação.

Propósito: Ajudar na priorização de recursos para educação básica.

4. A produção física industrial apresenta tendência de crescimento ou queda no período observado?

Hipótese: Há desaceleração em alguns setores associada a ciclos econômicos.

Propósito: Identificar segmentos industriais que exigem estímulos de competitividade.

5. Quais órgãos federais concentram o maior número de servidores ativos?

Hipótese: Defesa e Educação concentram a maior parte.

Propósito: Avaliar distribuição de força de trabalho e possíveis gargalos administrativos.

6. Há relação entre o tamanho do órgão e o número de pensionistas/inativos?

Hipótese: Órgãos historicamente antigos têm proporção maior de inativos.

Propósito: Auxiliar estudos sobre impacto previdenciário no setor público.

7. Estados com maior IDHM tendem a ter maior participação industrial?

Hipótese: Sim, indicando que o desenvolvimento humano acompanha a industrialização.

Propósito: Reforçar integração entre políticas industriais e sociais.

8. Como os indicadores do IDHM podem influenciar políticas de acesso ao ensino superior (ex.: UFC)?

Hipótese: Regiões com IDHM mais baixo demandam mais políticas afirmativas.

Propósito: Orientar distribuição de vagas e programas de apoio estudantil.

3. Dados Utilizados

A análise utilizou bases heterogêneas, com diferentes granularidades e formatos.

- Atlas do Desenvolvimento Humano: 261 colunas com indicadores sociais.

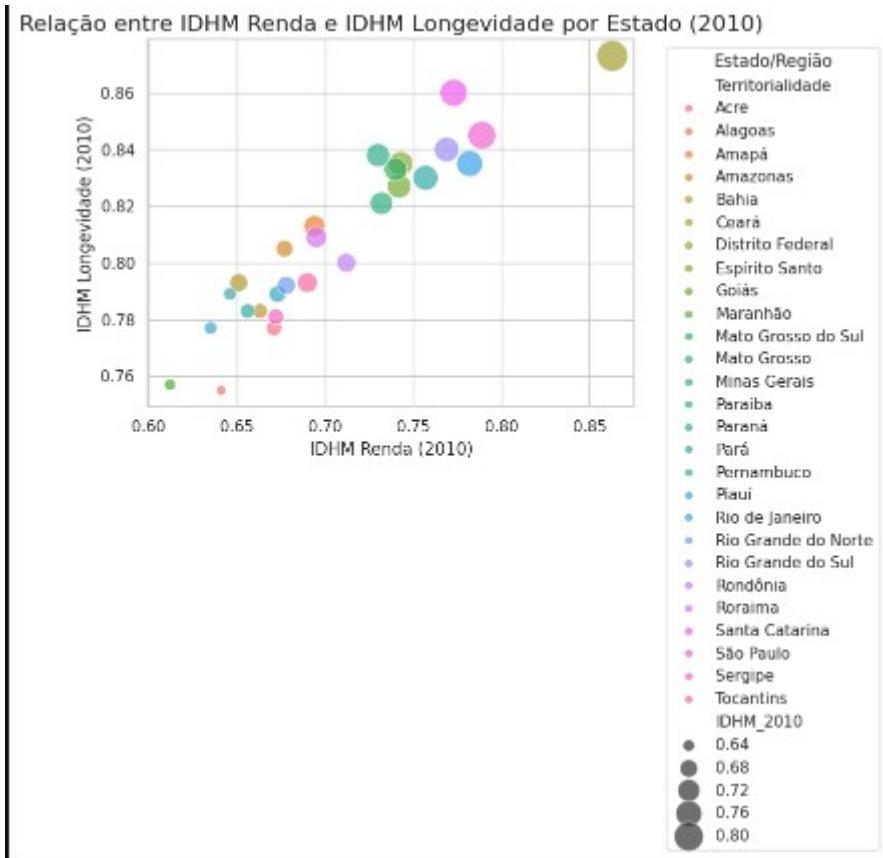
- PIMP: Séries numéricas com variações da produção industrial.

- Servidores por Órgão: 327 registros de repartições públicas.

Os dados foram tratados para remoção de valores nulos, padronização e seleção de colunas relevantes.

4. Principais Resultados

Evolução do IDHM (1991–2010)



Achado: Todas as unidades federativas apresentaram evolução significativa no IDHM. Estados como São Paulo, Santa Catarina e Distrito Federal apresentam valores mais altos, enquanto Maranhão e Alagoas mantêm os menores índices.

Interpretação: Evidencia melhora geral, mas reforça desigualdades históricas regionais.

Comparação IDHM Geral × IDHM Renda

- Achado: A dimensão Renda cresce mais lentamente que o IDHM geral na maioria dos estados.
- Aplicação: Políticas sociais podem priorizar renda, pois é o principal gargalo do índice.

Produção Industrial – Tendência Geral

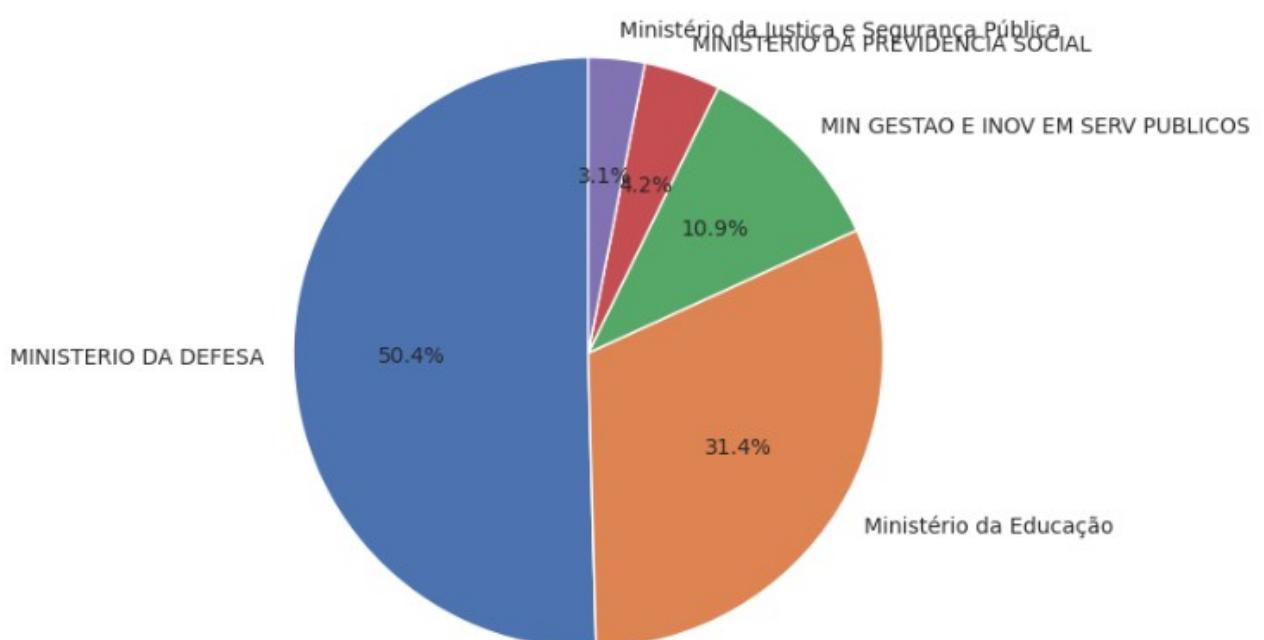


Achado: Observou-se forte oscilação na série, com períodos de queda acentuada que possivelmente refletem crises econômicas nacionais.

Aplicação: Setores vulneráveis precisam de políticas de estímulo e modernização tecnológica.

Relação Servidores Ativos × Pensionistas/Inativos

Distribuição de Servidores (Ativos/Inativos/Pensionistas) nos 5 Maiores Órgãos



Achado: Órgãos mais antigos (como Forças Armadas) concentram mais inativos e pensionistas.

Aplicação: Impacto direto na previdência pública e na distribuição orçamentária.

5. Limitações

- Estruturas heterogêneas dos datasets dificultaram análises integradas.
- A base do Atlas possui muitas colunas com dados faltantes.
- Falta de séries temporais completas para todos os indicadores industriais.

6. Recomendações

- Criar um processo de padronização das bases para facilitar análise integrada.
- Complementar bases do IDHM com dados mais recentes do IBGE (2022).
- Realizar análises preditivas de produção industrial para cenários futuros.
- Investir em dashboards interativos para tomada de decisões em tempo real.

7. Desafios e Aprendizados

Durante o desenvolvimento do trabalho, os principais desafios foram:

- Tratar arquivos com grande amplitude de colunas (Atlas);
- Corrigir inconsistências de encoding e delimitadores no arquivo CSV;
- Lidar com diferentes granularidades (setorial, estadual e institucional);
- Definir perguntas que fossem coerentes entre si apesar da diversidade dos datasets.

Como aprendizados, destacam-se:

- Maior domínio no uso de Python para exploração e visualização;
- Aprimoramento da habilidade de formular hipóteses testáveis;
- Importância da limpeza e transformação de dados antes da análise;
- Necessidade de interpretar resultados dentro de um contexto social e econômico.

8. Conclusão

O estudo permitiu identificar padrões relevantes em três grandes áreas: desenvolvimento humano, produção industrial e estrutura do serviço público. A análise mostrou crescimento contínuo do IDHM, fragilidades na dimensão renda, oscilações significativas na indústria e desigual distribuição de servidores entre os órgãos federais.

Os insights obtidos oferecem base sólida para decisões de políticas públicas, como alocação de recursos, definição de prioridades em educação e estratégias de fortalecimento industrial.

O trabalho também ressaltou a importância da integração de dados e da análise exploratória como etapa fundamental na ciência de dados.