

## DESAFIO I

### Ciência e Governança de Dados

#### Contextualização

O Brasil, um país de dimensões continentais e uma diversidade social, econômica e ambiental, é influenciado por múltiplos fatores e agentes que atuam de modo interligado para determinar os impactos socioeconômicos em diferentes regiões do país. A qualidade e cobertura dos serviços de saúde influenciam diretamente a expectativa de vida, o bem-estar e a produtividade; a educação repercute tanto na formação de cidadãos mais capacitados e conscientes quanto no potencial de crescimento econômico e na diminuição das desigualdades sociais. Questões ligadas à segurança pública, como violência, taxas de homicídio e criminalidade, afetam o bem-estar coletivo, a economia local e as oportunidades de desenvolvimento, especialmente em áreas urbanas e vulneráveis. Além disso, fenômenos ambientais (como desmatamento, secas, enchentes e poluição) provocam consequências econômicas e sociais significativas, afetando famílias, comunidades e setores produtivos. Fatores econômicos como desemprego, inflação e desigualdade de renda também contribuem para a instabilidade e perpetuação da pobreza e da vulnerabilidade social. Assim, avaliar e prever os agentes e fenômenos que mais causam impactos socioeconômicos no Brasil exige o uso integrado de dados multidimensionais para fundamentar políticas públicas mais eficazes, justas e voltadas às especificidades regionais e populacionais. Modelos preditivos, fundamentados em análises estatísticas e exploratórias, podem contribuir para identificar padrões, antecipar impactos, mapear regiões vulneráveis e direcionar decisões estratégicas voltadas à redução sustentável da pobreza e à promoção do desenvolvimento humano e social.

#### Pergunta

Como poderíamos avaliar e prever/visualizar os agentes/fenômenos que mais causam impactos socioeconômicos no Brasil?

Nesta primeira etapa, o objetivo é adquirir, organizar, explorar e visualizar os dados necessários para responder à pergunta proposta no desafio.

## Conjunto de Dados

Você deverá acessar e reunir diversas bases de dados brasileiras relevantes que incluem informações capazes de avaliar as condições socioeconômicas e seus determinantes multifatoriais.

**Observação:** Caso as bases de dados sejam grandes demais para processamento e análises, fique à vontade para utilizar uma região, um estado, um conjunto de municípios, ou mesmo, um município específico.

## Atividades que deverão ser desenvolvidas para responder à pergunta e demonstradas nos entregáveis

### Análise Exploratória e Preparação dos Dados

- Realizar a limpeza e tratamento dos dados: remover inconsistências, lidar com valores ausentes, transformar variáveis caso seja necessário, etc.
- Executar análise estatística inicial e identificação de padrões, correlações e tendências relevantes.
- Desenvolver visualizações iniciais que facilitem a compreensão das variáveis mais significativas e dos principais *insights* da análise exploratória.

### Entregáveis

- **Notebook Jupyter:** Deve conter todos os códigos utilizados na aquisição e exploração dos dados, acompanhados de comentários explicativos.
- **Arquivo README:** O README do repositório deve explicar a escolha dos dados, a metodologia de aquisição, os principais passos do notebook e os principais insights da análise exploratória.
- **Repositório:** A entrega deverá ser feita em um repositório público (e.g., Github ou similar) por meio do envio do link do repositório, permitindo fácil acesso, reprodutibilidade e versionamento do trabalho.
- **Visualizações dos dados:** É necessário produzir visualizações iniciais dos dados que permitam a compreensão dos principais padrões, *insights* e variáveis relevantes. As visualizações podem ser construídas utilizando qualquer tecnologia apropriada.

Critérios que serão utilizados para a avaliação

- Clareza na escolha e aquisição dos dados.
- Qualidade da análise exploratória e tratamento dos dados.
- Clareza no código, organização e explicação dos processos no notebook Jupyter.
- Utilidade, qualidade e clareza das visualizações produzidas.
- Habilidade de escrita e estrutura do relatório preliminar (README).
- Qualidade, reprodutibilidade e fidelidade dos entregáveis.
- Organização do repositório (Github).