

Projeto 4: Investigando Tiroteios Fatais Envolvendo Policiais nos EUA

Contextualização do Dataset

O conjunto de dados "**Fatal Police Shootings in the US**", compilado pelo *The Washington Post*, registra todos os tiroteios fatais envolvendo policiais em serviço nos Estados Unidos desde 2015. Esses dados são críticos para entender padrões de violência policial, disparidades raciais e fatores associados a esses incidentes, como uso de armas, saúde mental das vítimas e contexto geográfico.

Fonte dos Dados:

- Coletados pelo *The Washington Post* a partir de relatórios oficiais, notícias locais e bancos de dados independentes.
- Inclui informações como idade, raça, gênero das vítimas, tipo de arma utilizada, sinais de doença mental e presença de câmeras corporais.

Relevância Social:

- Esses dados ganharam destaque após protestos como o **Black Lives Matter**, destacando a necessidade de transparência e responsabilidade policial.
-

Situação Problema

Você foi contratado(a) como analista de dados por uma organização de direitos civis para investigar padrões em tiroteios fatais envolvendo policiais. O dataset fornecido está corrompido devido a problemas de codificação de caracteres, impossibilitando sua análise inicial. Além disso, há valores ausentes e inconsistências que precisam ser tratados.

Objetivo:

1. Corrigir a codificação do dataset para permitir seu carregamento.
 2. Realizar uma **Análise Exploratória de Dados (EDA)** para responder às seguintes perguntas:
 - Qual a distribuição de vítimas por raça e gênero?
 - Qual a porcentagem de casos onde havia sinais de doença mental?
 - Quais são os tipos de arma mais comuns nesses incidentes?
 - Existem padrões geográficos (estados/cidades) com maior incidência?
 - Qual a relação entre o nível de ameaça (*threat_level*) e o desfecho fatal?
-

Passos para a Solução

Passo 1: Corrigir a Codificação de Caracteres

- **Desafio:** O dataset não carrega devido a erros de codificação.

- **Ação:**

1. Use a biblioteca chardet para detectar a codificação correta do arquivo PoliceKillingsUS.csv.
2. Carregue o dataset usando a codificação identificada.

Passo 2: Limpeza e Preparação dos Dados

- **Desafio:** Dados ausentes e colunas inconsistentes.

- **Ações:**

1. Identifique e quantifique valores ausentes (ex: armed, age, race).
2. Decida como tratar os missing values (ex: remoção, imputação ou categorização como "desconhecido").
3. Padronize colunas categóricas (ex: race com valores como W, B, H).

Passo 3: Análise Exploratória (EDA)

- **Desafio:** Extrair insights relevantes para a organização.

- **Ações:**

1. Calcule estatísticas descritivas (média de idade, proporção por gênero).
2. Crie visualizações para:
 - Distribuição de vítimas por raça (gráfico de barras).
 - Proporção de casos com sinais de doença mental (gráfico de pizza).
 - Frequência de tipos de arma (gráfico de treemap ou histograma).
 - Mapa geográfico (opcional) destacando estados com mais incidentes.
3. Analise a correlação entre variáveis (ex: threat_level vs. signs_of_mental_illness).

Passo 4: Salvar o Dataset em UTF-8

- **Ação:** Exporte o dataset limpo em formato UTF-8 para uso futuro.

Entregáveis Esperados

1. **Notebook (.ipynb)** contendo:

- Código para correção de codificação e tratamento de dados.
- Visualizações e análises respondendo às perguntas do problema.
- Conclusões escritas sobre os padrões identificados.

2. **Dataset limpo** em formato CSV (UTF-8).

Dica: Utilize bibliotecas como pandas, matplotlib, seaborn e chardet. Documente cada decisão técnica (ex: por que escolheu remover ou imputar dados ausentes?).