**Proposta do projeto final de Data Science**

**Tema:** Violência Armada e suas tendências futuras nos EUA.

**Base:** Nossa base de estudo foi coletada no repositório online Kaggle para aprendizagem de máquina: [Dados de violência armada. Registro abrangente de mais de 260 mil incidentes de violência armada nos EUA entre 2013-2018](https://www.kaggle.com/jameslko/gun-violence-data)

**Informações sobre a base**

Registrou-se mais de 260 mil incidentes de violência armada, com informações detalhadas sobre cada incidente, disponíveis no formato CSV. O arquivo CSV contém dados de todos os incidentes registrados de violência armada nos EUA entre janeiro de 2013 e março de 2018, inclusive.

**Coleta dos dados:**

Fase 1: para cada data entre 1/1/2013 e 31/03/2018, um script Python consultou todos os incidentes que ocorreram naquela data em particular, depois digitalizou os dados e os escreveu em um arquivo CSV. Cada mês tem seu próprio arquivo CSV, com exceção de 2013, já que não foram registrados muitos incidentes a partir de então.

Fase 2: cada entrada foi aumentada com dados adicionais que não podem ser visualizados diretamente na página de resultados da consulta, como informações do participante, dados de geolocalização etc.

Fase 3: as entradas foram classificadas em ordem crescente e depois mescladas em um único arquivo CSV.

**Proposta**

Atualmente, faltam quantidades grandes e facilmente acessíveis de dados detalhados sobre a violência armada em geral.

Para tal resolvemos focar na análise do contexto Norte Americano por ser um país que permite a regulamentação do uso de armas em alguns estados, gerando assim um índice de violência que demanda uma maior incidência de crimes cometidos com o uso delas.

Como mostrado na descrição da base estudada, relatórios como [este](https://heavy.com/news/2018/02/nikolas-cruz-florida-instagram-nicolas-nicholas-social-media/) e [este](https://www.nytimes.com/2018/02/17/us/nikolas-cruz-florida-shooting.html) mostram que Nikolas Cruz exibiu muitos sinais de alerta nas redes sociais antes do tiroteio, por exemplo.

Para esse fim, a partir dessa base de dados, podemos identificar e levantar hipóteses que nos levem a entender melhor a necessidade de um reforço em tais áreas com maior índice de ocorrências desse viés, tais quais:

* Em que cidades ou condados têm maior probabilidade de ocorrer um crime à mão armada?
* Qual faixa etária de pessoas está presente na participação de tais crimes?
* Que tipo de armas possuem maior incidência nos casos de violência?
* Relacionar em quais distritos de votação para senadores ocorrem mais violência armada.