

# **Analysing Patterns and Risk Factors of Gun Violence in the US**

## **Data Science Project**

### **Dataset Description Document**

| <b>Name</b>       | <b>Roll Number</b> |
|-------------------|--------------------|
| Shaaf Salman      | 21L-6083           |
| Syed Farhan Jafri | 21L-6074           |
| Haider Khan       | 21L-6067           |

### **Problem Statement**

Gun violence remains a critical public safety issue in the United States, impacting communities nationwide. Understanding the patterns, trends, and contributing factors of such incidents is essential for developing effective prevention strategies and policy interventions. This project aims to analyse a comprehensive dataset of over 260,000-gun violence incidents recorded between January 2013 and March 2018. By leveraging data science techniques, we seek to identify key trends, geographic hotspots, demographic characteristics of victims and perpetrators, and potential risk factors associated with gun-related incidents. Our findings will provide valuable insights that can inform policymakers, law enforcement agencies, and public health officials in their efforts to mitigate gun violence and enhance community safety.

### **Source of the Dataset**

This dataset is obtained from Kaggle, it contains data downloaded from the Gun Violence Archive through web scraping techniques.

<https://www.kaggle.com/datasets/jameslko/gun-violence-data>

<https://www.gunviolencearchive.org/>

### **Description of the Dataset**

This dataset is a comprehensive record of over 260,000 US gun violence incidents recorded between January 2013 and March 2018. This dataset contains detailed information about each incident, available in CSV format. It contains date of incident, location (state, city, address), number of people killed/injured and other necessary data discussed ahead in the columns section. There are a total of 29 columns in this dataset.

## **Description of the Columns of the Dataset**

1. **incident\_id** – Unique identifier for each gun violence incident.
2. **date** – Date of the incident.
3. **state** – U.S. state where the incident occurred.
4. **city\_or\_county** – City or county where the incident occurred.
5. **address** – Specific location of the incident (if available).
6. **n\_killed** – Number of people killed in the incident.
7. **n\_injured** – Number of people injured in the incident.
8. **incident\_url** – Link to the incident details on the Gun Violence Archive website.
9. **source\_url** – News source reporting the incident.
10. **incident\_url\_fields\_missing** – Boolean indicating missing fields in the incident record.
11. **congressional\_district** – U.S. congressional district where the incident occurred.
12. **gun\_stolen** – Information on whether the gun was reported stolen.
13. **gun\_type** – Type(s) of gun(s) involved in the incident.
14. **incident\_characteristics** – Keywords describing the nature of the incident (e.g., mass shooting, domestic violence).
15. **latitude** – Latitude coordinate of the incident location.
16. **longitude** – Longitude coordinate of the incident location.
17. **location\_description** – Additional details about the location.
18. **n\_guns\_involved** – Number of guns involved in the incident.
19. **notes** – Additional notes or descriptions related to the incident.
20. **participant\_age** – Age(s) of individuals involved in the incident.
21. **participant\_age\_group** – Age group classification of participants (e.g., Adult 18+).
22. **participant\_gender** – Gender of individuals involved.
23. **participant\_name** – Names of individuals involved (if available).
24. **participant\_relationship** – Relationship between participants (if documented).
25. **participant\_status** – Status of participants (e.g., injured, killed, arrested).
26. **participant\_type** – Role of participants (e.g., victim, suspect).
27. **sources** – Additional sources reporting the incident.
28. **state\_house\_district** – State house legislative district where the incident occurred.
29. **state\_senate\_district** – State senate legislative district where the incident occurred.