

UNIVERSIDAD DE CASTILLA-LA MANCHA

MINERÍA DE DATOS

ENTREGABLE 1: DEFINICIÓN DEL PROBLEMA

---

## **Análisis de violencia con armas en EE. UU.**

---

DARÍO ANDRÉS FALLAVOLLITA  
FIGUEROA

ISRAEL MATEOS APARICIO RUIZ  
SANTA QUITERIA

FERNANDO POTENCIANO  
SANTIAGO

ADRIÁN JULIÁN RAMOS ROMERO

IGNACIO ROZAS LÓPEZ

LAURENTIU GHEORGHE ZLATAR

27 de octubre de 2023

## **Índice**

<b>1. Introducción y porcentajes de participación</b>	<b>2</b>
<b>2. Descripción breve de los datos originales</b>	<b>2</b>
<b>3. Antecedentes o trabajos similares</b>	<b>5</b>
<b>4. Planteamiento de la hipótesis y de los objetivos a perseguir</b>	<b>6</b>
<b>5. Posibilidades de enriquecimiento de los datos</b>	<b>7</b>

## 1. Introducción y porcentajes de participación

El presente documento tiene como fin definir el problema que abordará el proyecto. Este es el análisis de los incidentes violentos con armas en EE. UU.

El reparto de la participación entre integrantes ha sido equitativo, y se muestra en la tabla 1.

Apellidos y nombre	Correo	Participación
Fallavollita Figueroa, Darío Andrés	DarioAndres.Fallavollita@alu.uclm.es	16,6 %
Mateos Aparicio Ruiz Santa Quiteria, Israel	Israel.Mateos@alu.uclm.es	16,6 %
Potenciano Santiago, Fernando	Fernando.Potenciano@alu.uclm.es	16,6 %
Ramos Romero, Adrián Julián	AdrianJulian.Ramos@alu.uclm.es	16,6 %
Rozas López, Ignacio	Ignacio.Rozas@alu.uclm.es	16,6 %
Zlatar, Laurentiu Gheorghe	LaurentiuGheorghe.Zlatar@alu.uclm.es	16,6 %

Cuadro 1: Porcentaje de participación de los integrantes

## 2. Descripción breve de los datos originales

Los datos originales que se usarán provienen de tres fuentes, dando lugar a 4 conjuntos de datos: uno con incidentes violentos con armas, otro con datos de pobreza y otros dos con datos acerca de leyes sobre armas de fuego. Estos conjuntos de datos se han obtenido mediante el uso de técnicas de *web scraping*, incluyendo sus fuentes en estos documentos.

A continuación se presentan los diccionarios de datos para cada uno de ellos a modo de descripción.

### Incidentes violentos con armas en EE. UU.

El conjunto de datos principal trata sobre incidentes violentos con armas en EE. UU., obtenido de <https://www.kaggle.com/datasets/jameslko/gun-violence-data>. Se muestra en la tabla 2.

Columna	Tipo de dato	Descripción
incident_id	int	Identificador del incidente
date	date	Fecha en la que se produjo el crimen
state	String	Estado en el que se produjo el crimen
city_or_county	String	Ciudad/condado del crimen
address	String	Dirección del lugar del crimen
n_killed	int	Número de personas fallecidas
n_injured	int	Número de personas heridas
incident_url	url	URL acerca del incidente
source_url	String	Referencia a la fuente del informe
incident_url_fields_missing	boolean	Validación del URL: verdadero si incident_url está presente, falso en caso contrario
congressional_district	int	Identificador del distrito electoral
gun_stolen	String	Estado de las armas involucradas en el crimen.
gun_type	String	Tipificación de las armas utilizadas en el crimen
incident_characteristics	String	Características del incidente
latitude	float	Distancia de latitud
location_description	String	Localización concreta del incidente
longitude	float	Distancia de longitud
n_guns_involved	int	Número de armas involucradas en el incidente
notes	int	Información adicional del crimen
participant_age	String	Edad exacta de los participantes en el momento del crimen
participant_age_group	String	Grupo de edad de los participantes en el momento del crimen
participant_gender	String	Género de los participantes
participant_name	String	Nombre de los participantes involucrados en el crimen
participant_relationship	String	Relación del participante con otros participantes
participant_status	String	Grado de daño hecho al participante
participant_type	String	Papel del involucrado en el crimen
sources	url	Fuente de los participantes
state_house_district	int	Distrito electoral de la cámara de representantes para la elección de un representante
state_senate_district	int	Distrito electoral para elección de un senador estatal en una legislatura

Cuadro 2: Diccionario de datos del conjunto de datos de incidentes violentos con armas en EE. UU.

### Datos de pobreza en EE. UU.

El conjunto de datos que trata acerca de datos de pobreza en EE. UU. se ha obtenido de <https://www.povertyusa.org/data>. Se muestra en la tabla 3.

Columna	Tipo de dato	Descripción
year	int	El año al que corresponde la información en el registro
state	String	El estado de Estados Unidos al que se refiere el registro
population	int	La población total en el estado y año especificados
in_poverty	int	El número de personas que se encuentran en situación de pobreza (ganan menos que el umbral de pobreza oficial del gobierno, que para una familia de cuatro miembros es de unos 25,700 dólares)
poverty_rate	float	La tasa de pobreza, expresada como un porcentaje, que representa la proporción de la población que se encuentra en situación de pobreza
median_household_income	float	La mediana de los ingresos de los hogares. Incluye los ingresos del cabeza de familia y de todas las demás personas de 15 años o más del hogar, estén o no emparentadas con el cabeza de familia
deep_poverty_rate	float	Porcentaje de personas que viven en un hogar con unos ingresos totales en efectivo inferiores al 50 % del umbral de pobreza establecido por la Oficina del Censo de Estados Unidos
median_rent	float	La mediana de los costos del alquiler y los servicios de una vivienda de dos dormitorios
unemployment_rate	float	Porcentaje de trabajadores desempleados que han buscado trabajo activamente y están actualmente disponibles para trabajar
without_health_insurance	float	Porcentaje de personas menores de 65 años y por debajo del 138 % del umbral de pobreza que no tenían seguro médico en ningún momento del año
supplemental_poverty_measure	float	La Medida de Pobreza Suplementaria (SPM) tiene en cuenta a las personas que caerían en la pobreza sin ciertas prestaciones no monetarias que les ayudan a superar el umbral de pobreza. Estas prestaciones incluyen la ayuda alimentaria y los subsidios de alquiler, así como las ayudas basadas en los impuestos, como el crédito fiscal por ingresos del trabajo (EITC)

Cuadro 3: Diccionario de datos del conjunto de datos de pobreza en EE. UU.

### Leyes para armas de fuego en EE. UU.

Los conjuntos de datos que tratan acerca de datos de leyes sobre armas de fuego en EE. UU. se han obtenido de <https://mail.statefirearmlaws.org/resources>.

En realidad, se trata de un único conjunto de datos, mostrado en la tabla 4, mientras que el mostrado en la tabla 5 se trata de un *codebook* acerca del primero. Sin embargo, este último contiene una categorización de las leyes, que nos puede ser útil para analizar la importancia de cada una de estas categorías por separado. Por tanto, se incluyen ambos.

Columna	Tipo de dato	Descripción
state	String	Estado de EE. UU. al que corresponde la información del registro
year	int	Año al que corresponde la información del registro
felony, invcommitment, ...	boolean	Ley concreta que puede o no estar vigente
lawtotal	int	Cantidad total de leyes vigentes

Cuadro 4: Diccionario de datos del conjunto de datos de leyes para armas de fuego en EE. UU.

Columna	Tipo de dato	Descripción
Category Code	int	Identificador de categoría de leyes
Category	String	Categoría de la ley
Sub-Category	String	Subcategoría de la ley
Variable Name	String	Nombre de la ley en el anterior conjunto de datos. Corresponden a las definidas anteriormente: felony, danger, ...
Brief Description of Provision	String	Descripción breve de la ley
Detailed Description of Provision	String	Descripción detallada de la ley
Coding Notes	String	Notas adicionales para asignar un valor (vigente o no vigente) a cada ley
Coding Instructions	String	Instrucciones de qué valor asignar (vigente o no vigente) a cada ley
Notes	String	Notas adicionales
Data Source and Attribution	String	Fuente de todos los datos y autores

Cuadro 5: Diccionario de datos del *codebook* del conjunto de datos de leyes para armas de fuego en EE. UU.

### 3. Antecedentes o trabajos similares

Uno de los estudios más elaborados que hemos podido encontrar es el estudio elaborado por Amnistía Internacional, donde denuncian que más de 600 personas mueren al día por consecuencia de la violencia con armas de fuego y su fácil acceso.

En el artículo que subieron a su página web (<https://www.amnesty.org/es/what-we-do/arms-control/gun-violence/>) también llegan a conclusiones como que las personas racializadas son uno de los grupos que, de forma desproporcionada, más sufren los efectos de la violencia con armas, así como las personas de comunidades desfavorecidas y mujeres en situación de violencia de género intrafamiliar.

Según los datos que Amnistía Internacional ha analizado, la gran mayoría de la violencia con armas de fuego ocurre fuera de los entornos de conflicto armado y un porcentaje importante de las muertes por arma de fuego (llegando a más de la mitad en algunos estados de Estados Unidos) son suicidios o accidentes. De todos los homicidios cometidos en el mundo, el 71 % son consecuencia de las armas de fuego, siendo el perfil de la víctima y a la vez agresor jóvenes varones. La mujer, en concreto dentro de una pareja, es especialmente vulnerable cuando la pareja tiene fácil acceso a armas.

De todas las armas que hay en circulación, el 85 % las tienen particulares, el 13 % militares y un 2 % son para cuerpos policiales.

La violencia con armas de fuego está concentrada en barrios urbanos, pobres, con alta delincuencia y con presencia de narcotráfico

Una relación interesante a la que llegan es el vínculo entre la discriminación y la violencia con armas de fuego. Según los datos del 2020 el 60,9 % de las víctimas mortales por arma de fuego en estados unidos eran afroamericanos, siendo estos tan solo el 13 % de la población. Dicha discriminación se hace evidente en los jóvenes varones negros de entre 15 y 34 años, que tienen hasta 10 veces más probabilidades de morir por armas de fuego que su contraparte blanca.

Otro grupo preocupante son los infantes, que en estados unidos las armas de fuego es la primera causa de muerte infantil y entre un 2 y el 7 % de todas las heridas tratadas

Estados Unidos es un caso particular por tener cifras significativamente más altas que otros países industrializados. No tienen registro nacional de armas y se puede portar armas en público y solo tres estados y el Distrito de Columbia prohíben portar armas a la vista en público. Lidera en número de tiroteos masivos al año y la cifra no para de aumentar año tras año

Para terminar, se proponen algunas medidas para restringir el acceso a las armas de fuego argumentado que la gran mayoría de países del mundo tienen fuertes regulaciones al respecto y que los países con los controles más estrictos tienen una violencia con armas de fuego bastante baja como son Japón, Corea del Sur y Singapur. Un caso de éxito que tomaron estas medidas fue Australia, donde consiguió que las muertes por armas de fuego se redujeran a más de la mitad entre 1996 y 2019.

#### **4. Planteamiento de la hipótesis y de los objetivos a perseguir**

El principal objetivo del proyecto es el de analizar y comprender los patrones asociados a los incidentes violentos con armas en EE. UU. Para ello, pretendemos explorar las posibles relaciones entre la violencia armada y factores socioeconómicos en los estados de EE. UU (en concreto, con la pobreza), así como la influencia de la época en la que tienen lugar los incidentes.

Además, un componente esencial del análisis será el de relacionar estos incidentes con el ámbito jurídico, con el fin de determinar el impacto de las leyes de control de armas.

Nuestro objetivo final será el de presentar los resultados, sacar conclusiones y ofrecer recomendaciones para abordar la problemática de la violencia con armas en EE. UU.

Para concretar más estos objetivos, se ha planteado cuatro hipótesis:

1. Se producen más incidentes con armas los fines de semana.
2. Se producen más incidentes con armas en los meses de verano.
3. Los incidentes con armas son más frecuentes en los estados con mayor pobreza.
4. Cuantas más leyes sobre el uso de armas existen en un estado, menos incidentes con armas se producen en él.

## 5. Posibilidades de enriquecimiento de los datos

Dado que ya se han expuesto y descrito todos los datos que se usarán, en este apartado diferenciaremos los conjuntos de datos en principales o secundarios, y por las hipótesis en las que se utilizarán.

- **Incidentes violentos con armas en EE. UU.** Es el conjunto de datos principal, por lo que se utilizará para comprobar todas las hipótesis.
- **Datos de pobreza de EE. UU.** Conjunto de datos secundario. Se utilizará en conjunto con el principal para comprobar la tercera hipótesis (*los incidentes con armas son más frecuentes en los estados con mayor pobreza*).
- **Leyes para armas de fuego en EE. UU.** Conjunto de datos secundario. Se utilizará en conjunto con el principal para comprobar la cuarta hipótesis (*cuantas más leyes sobre el uso de armas existen en un estado, menos incidentes con armas se producen en él*).