

Análisis de violencia con armas en EE. UU.



Minería de Datos



Contenidos



Dominio del problema e hipótesis



Recogida de datos



Preproceso y transformación



Minería de datos



Resultados

Dominio del problema

Incidentes
de violencia
con armas

Estados
Unidos

2014 a 2017

1. Se producen más incidentes con armas los fines de semana.
2. Se producen más incidentes con armas en los estados con climas más cálidos.
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.

Hipótesis

Análisis extra

Análisis combinado con los
datos climáticos, de
pobreza y de leyes



Recogida de datos

Incidentes con
armas

Temperatura y
precipitaciones

Indicadores de
pobreza

Leyes de regulación
de armas vigentes

Poblaciones

Preproceso

Limpieza

Integración

Reducción

Transformación

Mayores problemas encontrados



VARIABLES DE
INCIDENTES MUY
DISTINTAS



DATOS CLIMÁTICOS
EN FICHEROS DE
TEXTO



INDICADORES DE
POBREZA MUY
CORRELACIONADOS



DIVISIÓN DE LEYES
POR CATEGORÍA



FUERTE
DEPENDENCIA CON
LA POBLACIÓN

Tarjetas de datos: hipótesis 1

| Columna | Tipo de dato | Descripción |
|---------------------|--------------|--|
| state | String | Estado al que se refiere el registro |
| year | int | Año al que se refiere el registro |
| is_weekend | int | 1 si se refiere a incidentes producidos en fin de semana, 0 en caso contrario |
| n_incidents_per_day | float | Número de incidentes con armas al día por cada 100.000 habitantes, 0 en caso contrario |

Tarjetas de datos: hipótesis 2

| Columna | Tipo de dato | Descripción |
|-----------------------|--------------|---|
| state | String | Estado al que se refiere el registro |
| year | int | Año al que se refiere el registro |
| month | int | Mes al que se refiere el registro |
| n_incidents | float | Número de incidentes por 100.000 habitantes |
| average_temperature | float | Temperatura media mensual en Celsius |
| average_precipitation | float | Precipitaciones medias mensuales en mm |

Tarjetas de datos: hipótesis 3

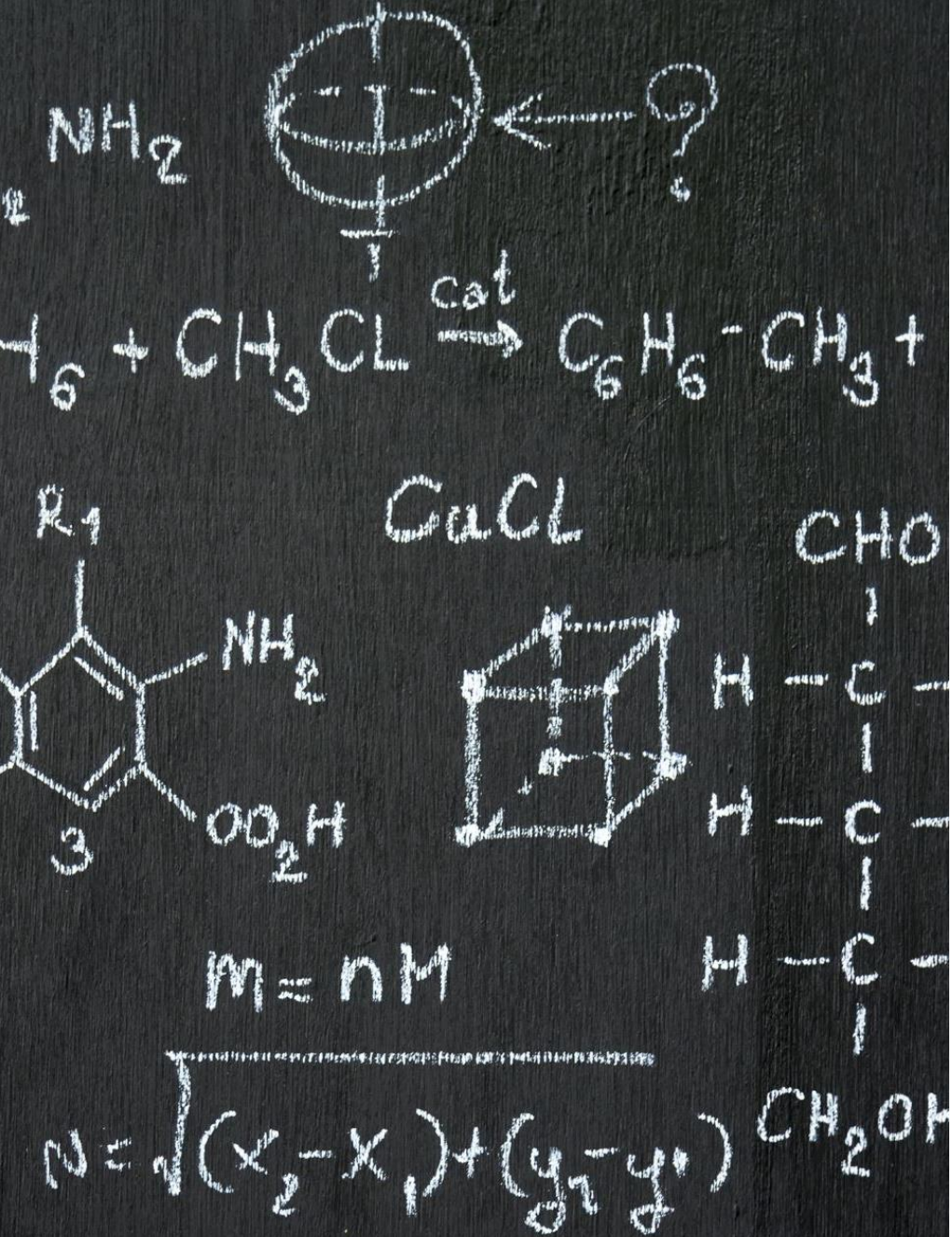
| Columna | Tipo de dato | Descripción |
|--------------|--------------|--|
| state | String | Estado al que se refiere el registro |
| year | int | Año al que se refiere el registro |
| n_incidents | float | Número de incidentes por 100.000 habitantes |
| poverty_rate | float | La tasa de pobreza, que representa el porcentaje de la población viviendo por debajo del umbral de pobreza (para una familia de 4 miembros, unos 27.500 dólares) |

Tarjetas de datos: hipótesis 4

| Columna | Tipo de dato | Descripción |
|-------------|--------------|---|
| state | String | Estado al que se refiere el registro |
| year | int | Año al que se refiere el registro |
| n_incidents | float | Número de incidentes por 100.000 habitantes |
| lawtotal | int | Cantidad total de leyes vigentes |
| laws_1 | int | Cantidad de leyes de categoría 1 vigentes |
| laws_2 | int | Cantidad de leyes de categoría 2 vigentes |
| ... | | |
| laws_14 | int | Cantidad de leyes de categoría 14 vigentes |

Tarjetas de datos: análisis combinado

| Columna | Tipo de dato | Descripción |
|-----------------------|--------------|---|
| state | String | Estado al que se refiere el registro |
| year | int | Año al que se refiere el registro |
| n_incidents | float | Número de incidentes por 100.000 habitantes |
| average_temperature | float | Temperatura media anual en Celsius |
| average_precipitation | float | Precipitaciones medias anuales en mm |
| poverty_rate | float | La tasa de pobreza, que representa el porcentaje de la población viviendo por debajo del umbral de pobreza (para una familia de 4 miembros, unos 27.500 dólares |
| lawtotal | int | Cantidad total de leyes vigentes |
| laws_1 | int | Cantidad de leyes de categoría 1 vigentes |
| laws_2 | int | Cantidad de leyes de categoría 2 vigentes |
| ... | | |
| laws_14 | int | Cantidad de leyes de categoría 14 vigentes |



Minería de Datos

H1: Regresión lineal

H2: Clustering con k-means++

H3: Regresión lineal

H4: Clustering con k-means++

Extra: Clustering con k-means++

Resultados: H1 (fines de semana)

MSE = 6,012e-4

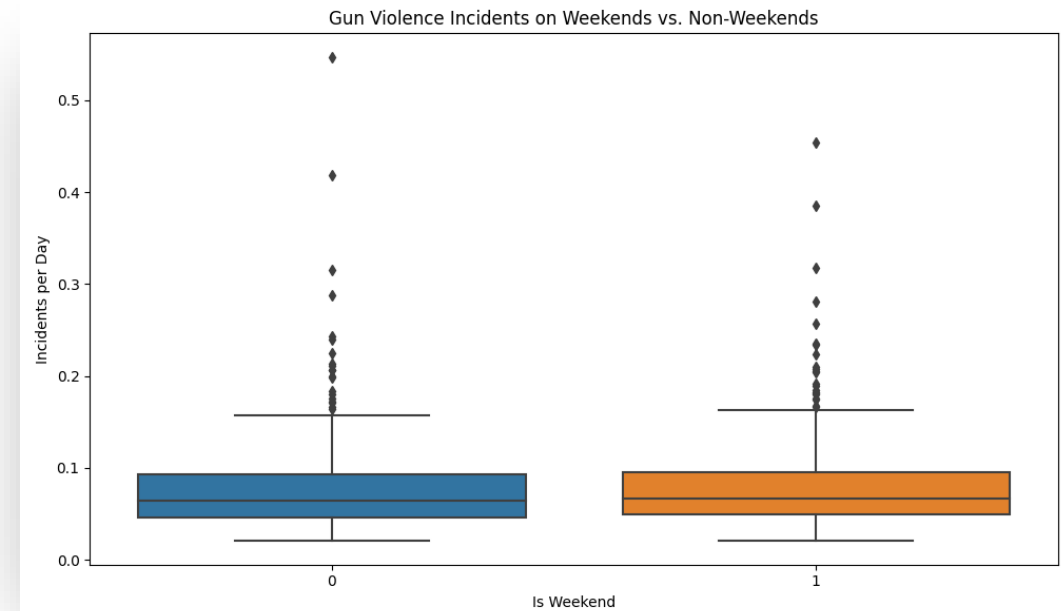
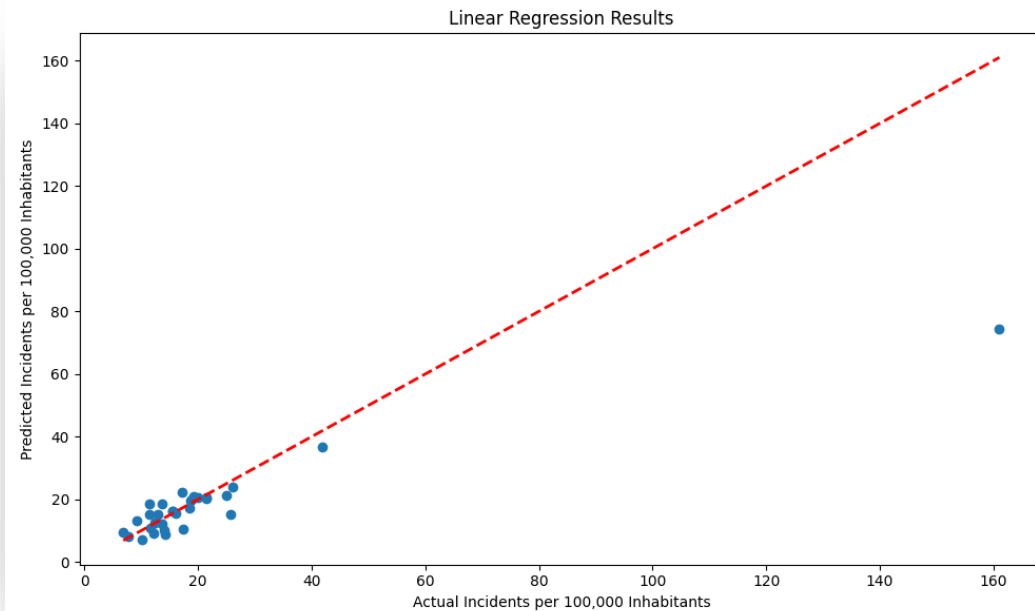
RMSE = 2,452e-2

MAE = 9,978e-3

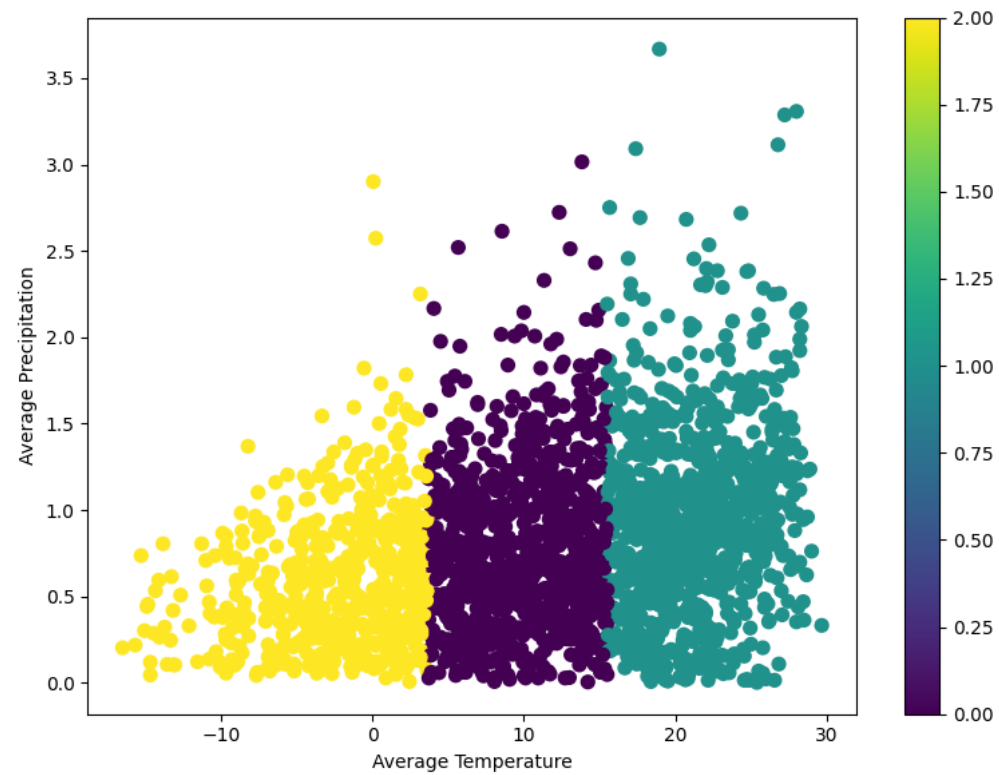
MAPE = 10,40%

R² = 88,15%

p-valor = 0,048



Resultados: H2 (clima)



| cluster | n_incidents | average_temperature | average_precipitation |
|---------|-------------|---------------------|-----------------------|
| 2 | 1.301505 | -2.401835 | 0.596606 |
| 0 | 1.519505 | 9.773176 | 0.795774 |
| 1 | 1.722896 | 21.194531 | 0.938084 |

Resultados: H3 (pobreza)

MAE = 255,33

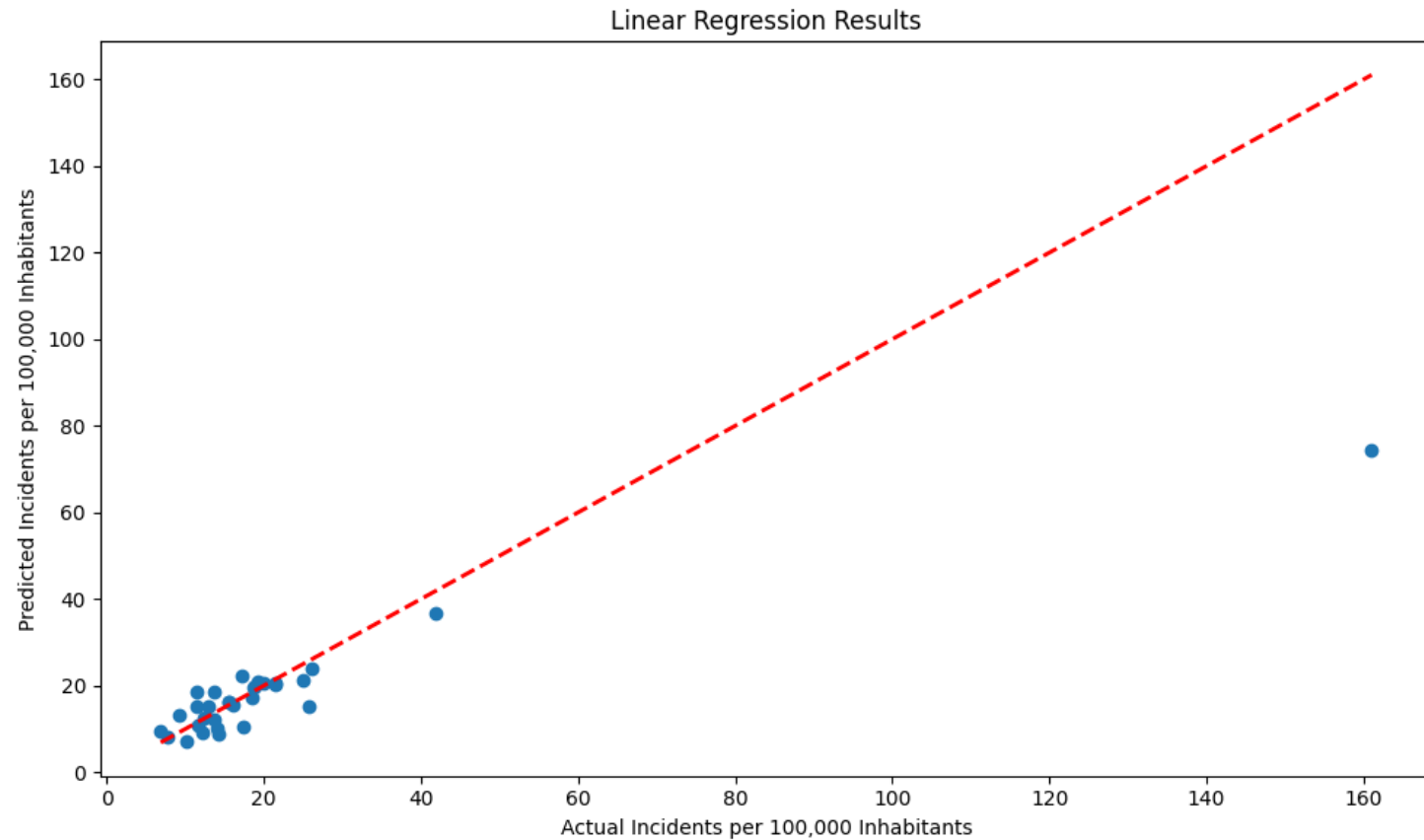
RMSE = 15,98

MAE = 5,561

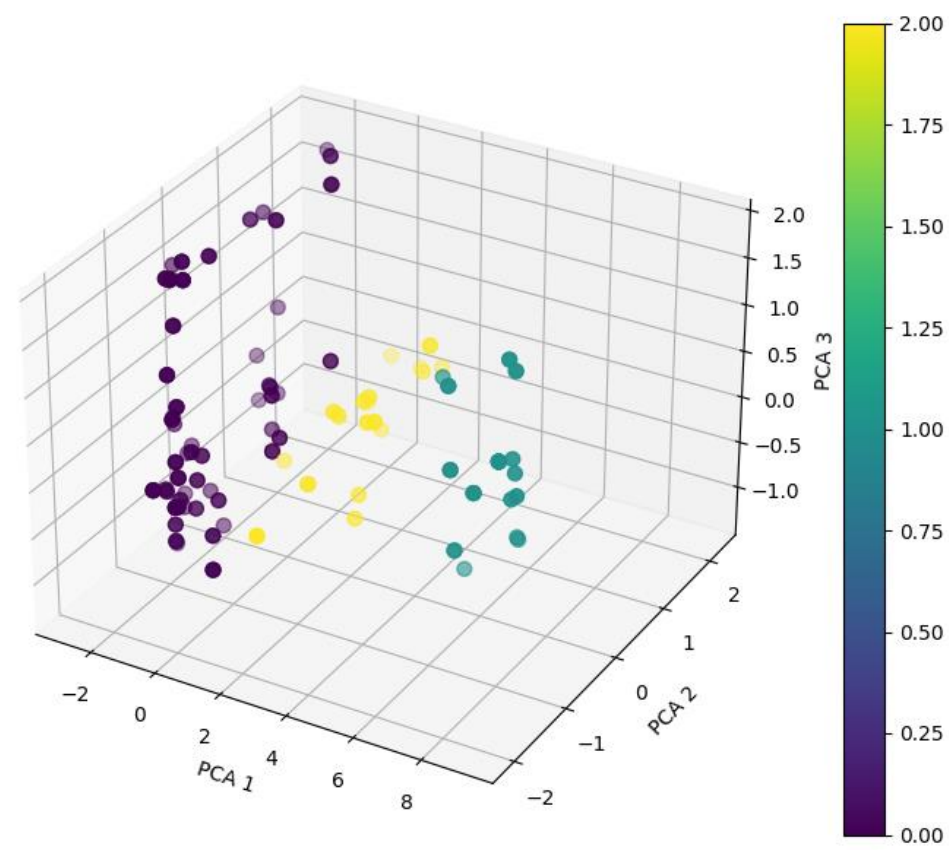
MAPE = 20,14%

$R^2 = 63,28\%$

p-valor = 0,857



Resultados: H4 (leyes)



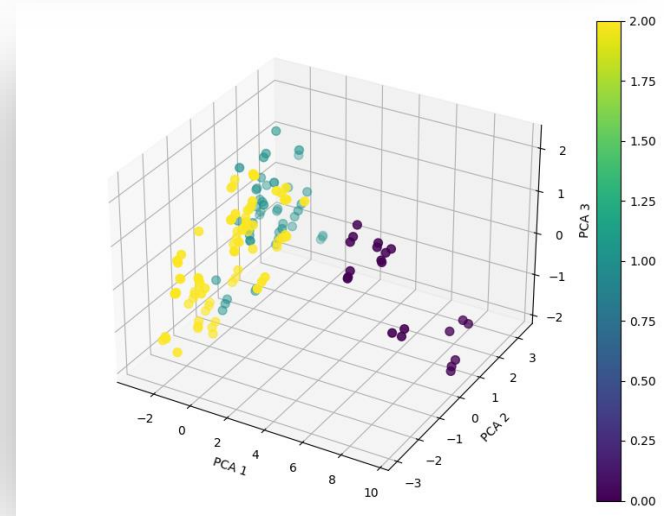
| cluster | n_incidents | lawtotal | laws_1 | laws_2 | laws_3 | laws_4 | laws_5 |
|---------|-------------|----------|--------|--------|--------|--------|--------|
| 1 | 16.35 | 79.97 | 9.12 | 9.50 | 7.38 | 7.75 | 3.25 |
| 2 | 17.54 | 33.35 | 4.20 | 2.83 | 3.57 | 5.75 | 0.50 |
| 0 | 19.05 | 11.36 | 0.62 | 0.56 | 2 | 0.27 | 0.12 |

Resultados: H4 (leyes)

| Variable | Desviación estándar |
|----------|---------------------|
| lawtotal | 35.033655 |
| laws_14 | 6.446223 |
| laws_2 | 4.646745 |
| laws_1 | 4.267830 |
| laws_4 | 3.875011 |
| laws_9 | 2.870923 |
| laws_3 | 2.763188 |
| laws_8 | 2.587734 |
| laws_6 | 2.244060 |
| laws_5 | 1.706300 |
| laws_7 | 1.291560 |
| laws_12 | 1.245325 |
| laws_10 | 1.163547 |
| laws_13 | 0.454162 |
| laws_11 | 0.375108 |

Resultados: análisis combinado

| cluster | n_incidents | average_temperature | average_precipitation | poverty_rate | lawtotal | laws_1 | laws_2 | laws_3 |
|---------|-------------|---------------------|-----------------------|--------------|----------|--------|--------|--------|
| 2 | 17.64 | 8.94 | 0.67 | 0.13 | 18.95 | 2.04 | 1.42 | 2.65 |
| 0 | 18.96 | 11.76 | 0.83 | 0.12 | 80.57 | 9.29 | 8.76 | 7 |
| 1 | 21.75 | 16.86 | 1.03 | 0.18 | 12.94 | 0.56 | 0.56 | 2 |



Conclusiones finales

01

No hay más
incidentes en
fines de semana

02

Hay más
incidentes en
climas más
cálidos

03

No hay más
incidentes en
estados más
pobres

04

Hay menos
incidentes en
estados con
mayor
regulación



Gracias por su atención. ¿Preguntas?

Darío Andrés Fallavollita Figueroa

Laurentiu Gherghe Zlatar

Israel Mateos Aparicio Ruiz Santa Quiteria

Fernando Potenciano Santiago

Adrián Julián Ramos Romero

Ignacio Rozas López

