

Por Camilo Rubiano Rafael Rincon Chafic Aljure

Gerencia de Proyectos en Ciencia de Datos



Acreditada en Alta Calidac Res. n°. 023654 del Mineducación.





Definición de Roles

Project Manajer: reuniones con stakeholders cumpliendo con los

tiempos de entrega

Data Scientist: Modelado, análisis avanzado y extracción de

insight

Data Analyst: Reportes, visualización y análisis exploratorio



La metodología OSEMN (Obtención, Limpieza, Exploración, Modelado y Narración)

Fase Inicial: Exploración de los Datos

Fase Intermedia: Análisis Profundo

Fase Avanzada: Modelado Predictivo

Impacto Potencial y Consideraciones Éticas



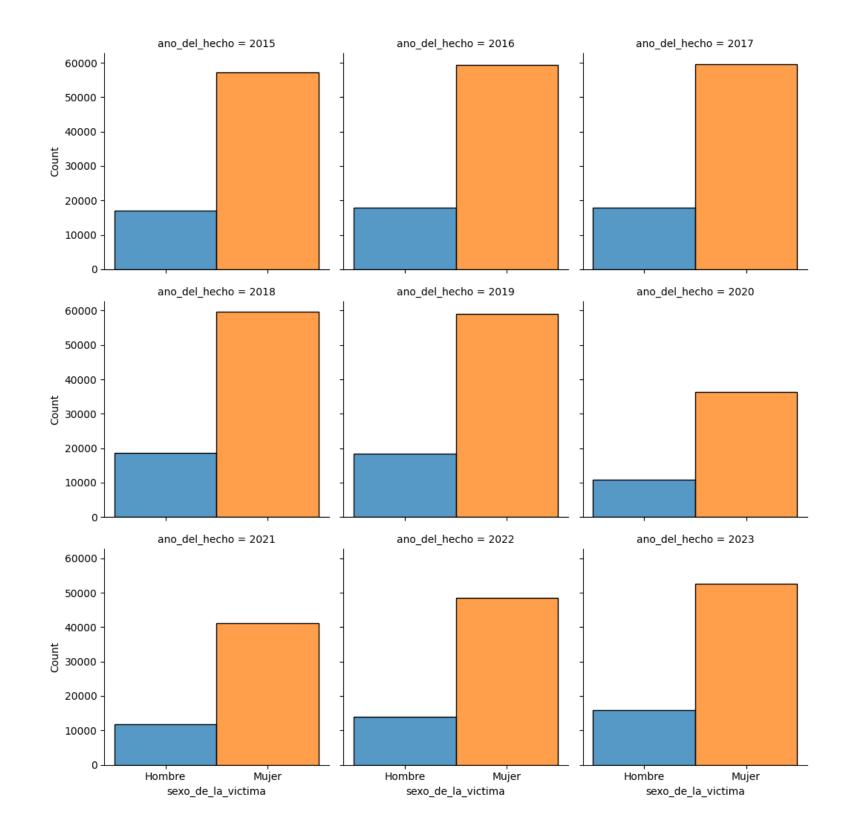
Data set

El data set usado contiene información sobre casos de violencia intrafamiliar ocurridos en Colombia durante los años 2015 y 2023. Los datos fueron suministrados por el instituto Nacional de Medicina Legal y ciencia Forense, cuenta con un total de 37 variables

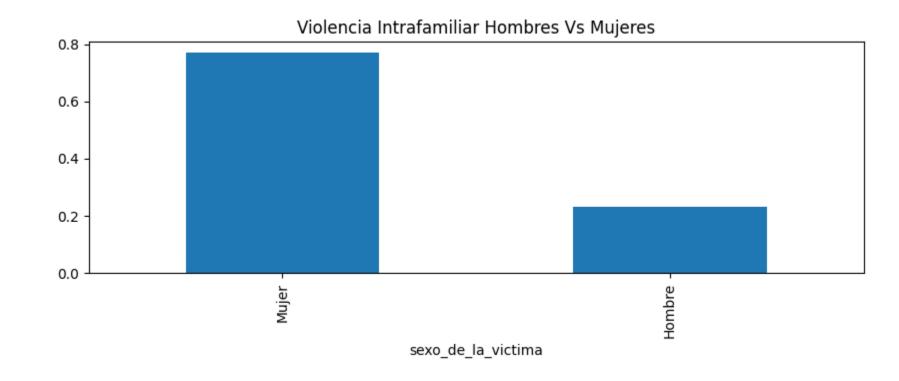




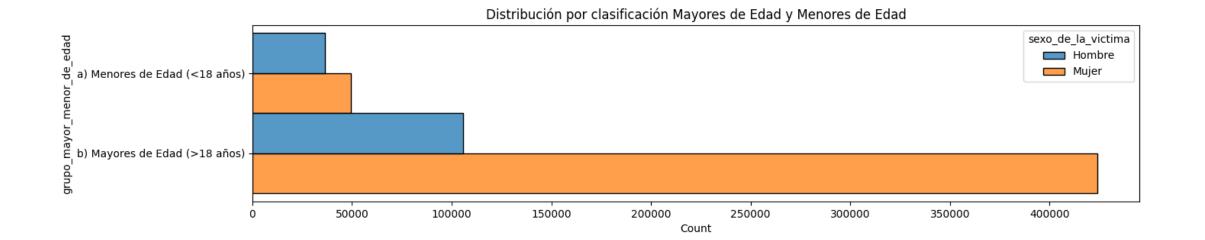
Estadística Descriptiva



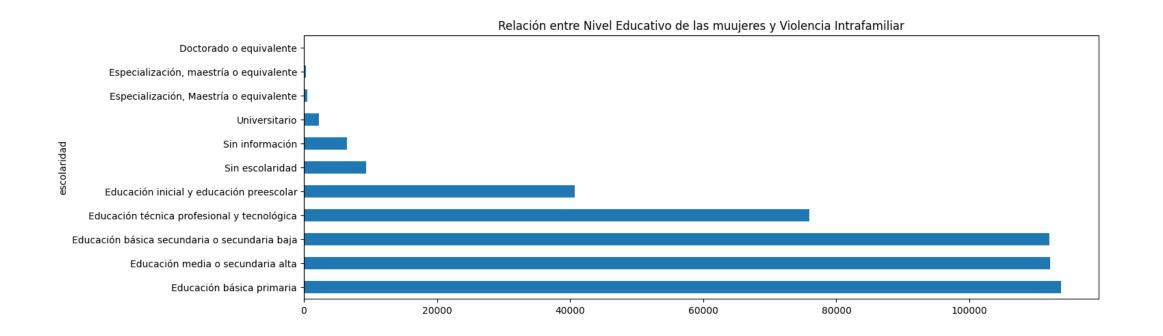


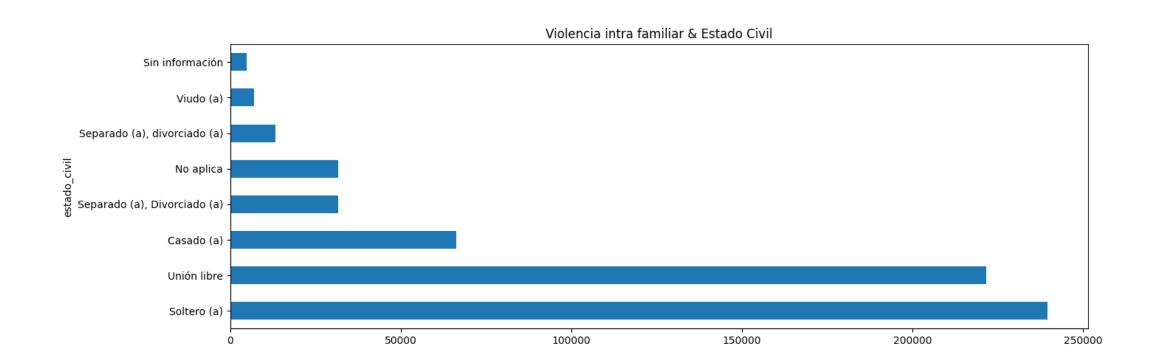


76% de los datos corresponden a mujeres afectadas contra un 23% de hombre

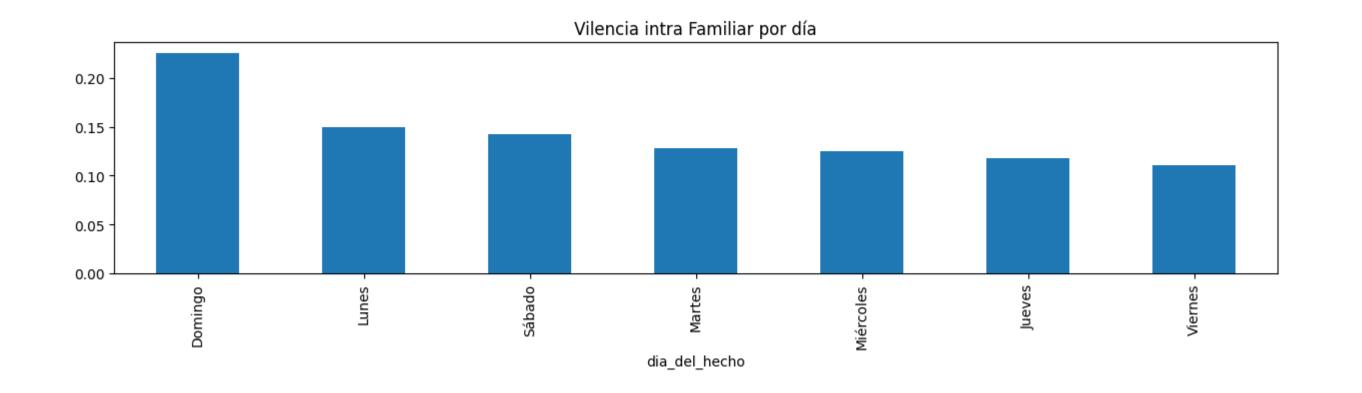


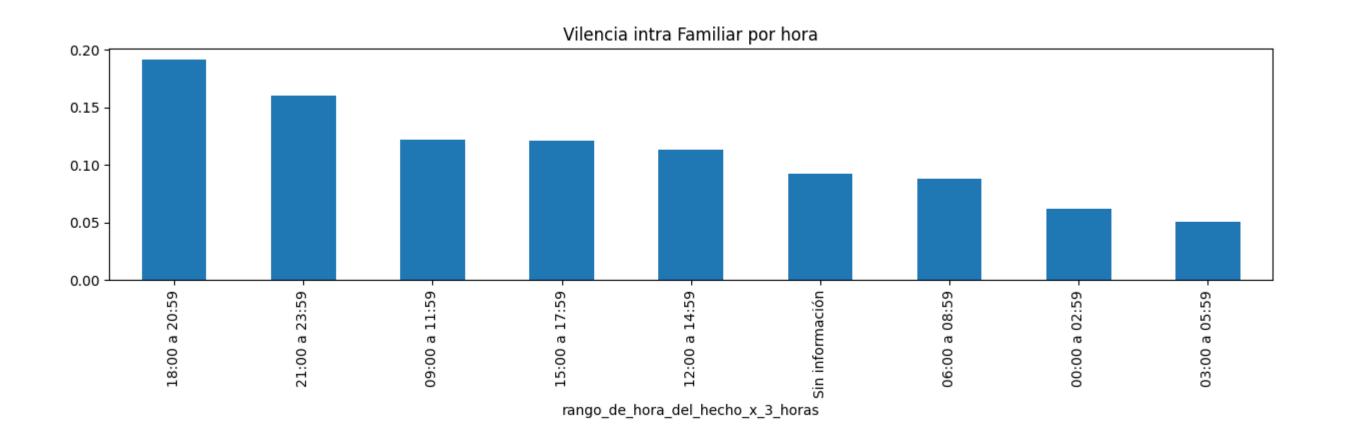






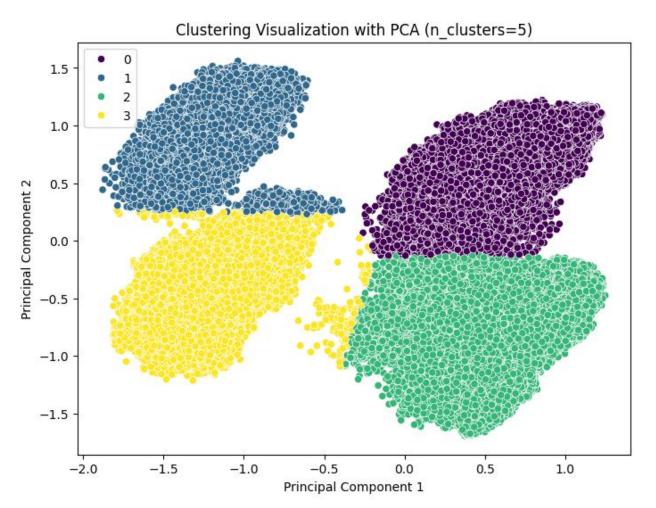


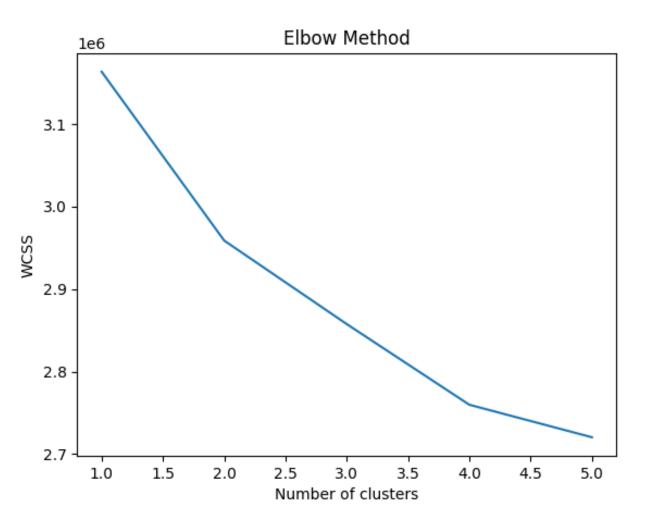






Transformación y clustering





Los clusters parecen estar diferenciados principalmente por: El sexo de la víctima (mujeres en su mayoría en los Clusters 1 y 2). La edad de la víctima (menores de edad en el Cluster 3). El tipo de agresor (padre y tío en los Clusters 0 y 3).

El Cluster 0 tiene una proporción moderada (64.43%).



Transformación y clustering

Cluster 0

Año del hecho: Las proporciones de los años están bastante equilibradas, con un ligero aumento en 2019 (12.92%) y 2023 (10.30%).

Sexo de la víctima: El 64.43% de las víctimas son mujeres.

Grupo de edad: El 100% de los casos corresponden a mayores de edad (>18 años). Presunto agresor: Los agresores más comunes son el padre (5.99%) y el tío (5.55%).

Cluster 1

Año del hecho: Similar al Cluster 0, con proporciones equilibradas y un ligero aumento en 2023 (12.33%).

Sexo de la víctima: El 83.47% de las víctimas son mujeres.

Grupo de edad: El 98.78% de los casos corresponden a mayores de edad (>18 años).

Presunto agresor: No hay un agresor predominante, ya que la mayoría de los valores son cercanos a 0.

Cluster 2

Año del hecho: Proporciones similares a los clusters anteriores, con un ligero aumento en 2016 (13.52%) y 2017 (13.72%).

Sexo de la víctima: El 90.15% de las víctimas son mujeres.

Grupo de edad: El 99.82% de los casos corresponden a mayores de edad (>18 años).

Presunto agresor: No hay un agresor predominante, con valores cercanos a 0.

Cluster 3

Año del hecho: Proporciones equilibradas, con un ligero aumento en 2019 (13.69%) y 2023 (11.55%).

Sexo de la víctima: El 55.53% de las víctimas son mujeres.

Grupo de edad: Solo el 18.63% de los casos corresponden a mayores de edad (>18 años), lo que indica que este cluster está más asociado a menores de edad.

Presunto agresor: Los agresores más comunes son el padre (23.95%) y el tío (4.21%).



Modelo y transformación

Regresión Logistica (Mejor Modelo)

•Exactitud: 0.41

Precision: 0.6816088079991012
Recall: 0.4123003737682637
F1-score: 0.5138042005420054

•Hamming Loss: 0.0267937358383135

RandomForestClassifier

•Exactitud: 0.26

Precision: 0.6594603653427182
Recall: 0.2674142031940197
F1-score: 0.3805241272604197

•Hamming Loss: 0.029897347462155773

GradientBoosterClasifier

•Exactitud: 0.20

Precision: 0.7292455210698966Recall: 0.1963982330954808F1-score: 0.30945497376592784

•Hamming Loss: 0.03009803212354708 a proporción muy alta de víctimas mujeres

(83.47% y 90.15%, respectivamente).



Propuestas

Mayor difusión en medios
Incentivar los valores en edades tempranas
Desnormalizando la violencia de genero
Incrementar los canales de denuncia y las campañas de prevención en días claves
Ampliar el acceso a la educación aprovechando las tecnologías de hoy
Leyes más estrictas contra los agresores
Campañas para identificar potenciales casos que aún no se denuncian



