

# MINERÍA DE DATOS APLICADA AL ANÁLISIS DE DELITOS EN LOS ÁNGELES (2020-2025)

PROYECTO FINAL - DATA MINING

**GRUPO  
3**

 Karen Torrico

 Karem Huacota

 Yesika Luna

 Elvis Miranda

 Ivan Mamani

Agosto, 2025

# CONTENT



01

Introducción

02

Definición del Problema

03

Problemas Resueltos

04

Métodos Utilizados

05

Resultados del Análisis

06

Conclusiones y  
Recomendaciones

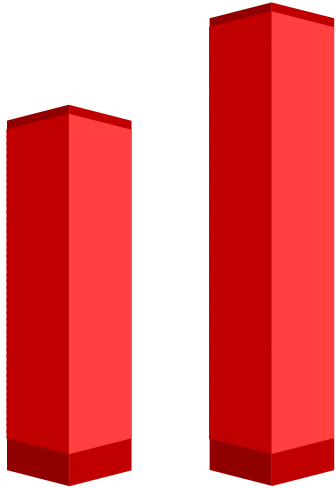
# Introducción

01



# Importancia del Análisis Delictivo

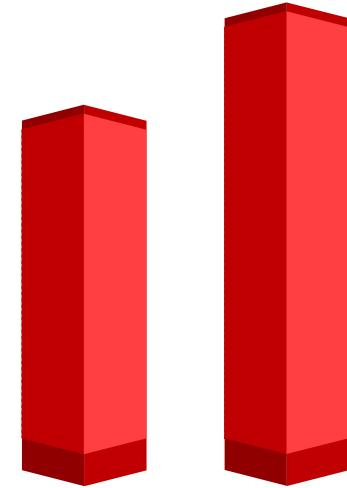
## Políticas Públicas y Seguridad



El análisis facilita la prevención, optimiza recursos policiales y mejora la confianza comunitaria.



## Enfoque en Los Ángeles



La ciudad es uno de los centros urbanos más grandes de EE.UU., con datos desde 2020 hasta 2025.





## Retos de Manejo y Procesamiento

Se requieren procesos rigurosos de limpieza y transformación para asegurar la calidad.

## Variables Clave

Fecha, hora, tipo de crimen, demografía de víctimas, ubicación geográfica y armas utilizadas.

## Amplio Dataset

Más de un millón de registros (1,004,991 filas) distribuidos en 28 columnas.

1

2

3



# Procesamiento de Datos

---

LOGO

## Limpieza y Transformación

Eliminación de datos incompletos, normalización y creación de nuevas variables.

## Aseguramiento de la Calidad

Es fundamental para asegurar la calidad y consistencia de la información.



# Resultados y Avances

LOGO

## Técnicas Aplicadas

Minería de datos, aprendizaje automático supervisado/no supervisado.



## Modelos Predictivos

Modelos con alta precisión para predecir la violencia en crímenes representan un avance.



LOGO

# Definición del Problema







# Objetivos del Análisis

---

LOGO

## **Análisis del Comportamiento Delictivo**

Analizar el comportamiento de los delitos reportados en Los Ángeles desde 2020 a 2025 y predecir si un crimen será violento o no.

1

## **Desarrollo de Modelos Predictivos**

Anticipar la violencia utilizando las variables disponibles en el dataset.

2



## **Tipos de Delitos Frecuentes**

¿Cuáles son los delitos más frecuentes en Los Ángeles durante el periodo?

*01*

## **Patrones Temporales y Geográficos**

¿Existen patrones temporales/geográficos en crímenes violentos (hora, día, zonas)?

*02*

## **Predicción de la Violencia**

¿Es posible predecir si un crimen es violento con la ubicación, hora y tipo de delito?

*03*



# Descripción del Dataset

---

LOGO

## Fuente de Datos

Dataset oficial "Crime Data from 2020 to Present" (Los Ángeles).

## Cobertura Temporal

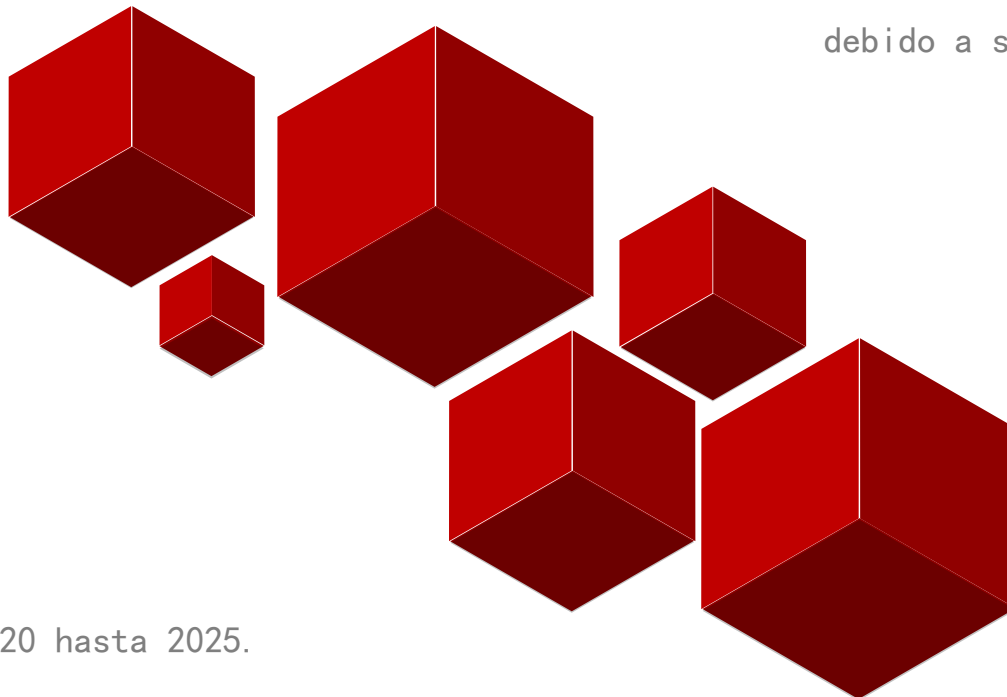
Incidentes delictivos reportados desde 2020 hasta 2025.

## Actualización Periódica

El dataset requiere limpieza y transformación debido a su tamaño y heterogeneidad.

## Variables Relevantes

Fechas, horas, ubicación, tipo de crimen, demografía de víctimas e información de armas.



LOGO

# Problemas Resueltos en los Datos



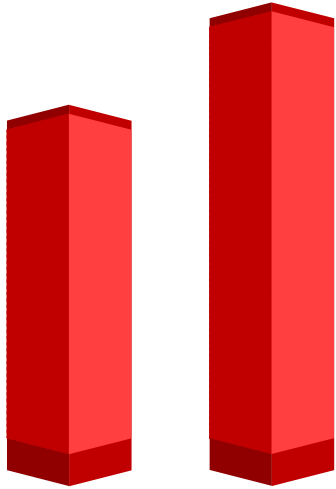
03



# Carga y Exploración Inicial de Datos

LOGO

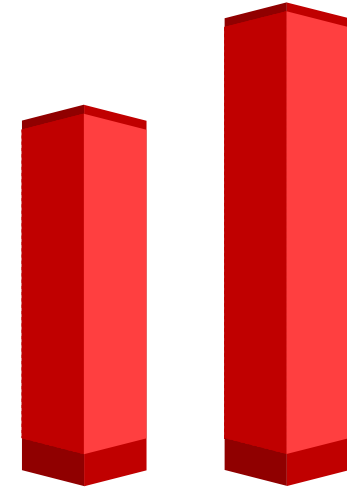
## Carga del Dataset



Carga de un gran dataset con más de un millón de registros y 28 columnas.



## Conocer la estructura inicial



Es necesario identificar la estructura y calidad de la información para planificar la limpieza.





## Identificación de Valores

### Nulos

Evaluar la cantidad de valores nulos en cada variable.



## Estrategias de Eliminación o Imputación

Para evitar sesgos o errores en el análisis, se requiere identificar y definir filtros.





# Transformación de Datos

LOGO

## Parseo de Fechas Frecuentes

Conversión de columnas de fechas a formato datetime. La conversión permite extraer información temporal.

## Parseo de Hora

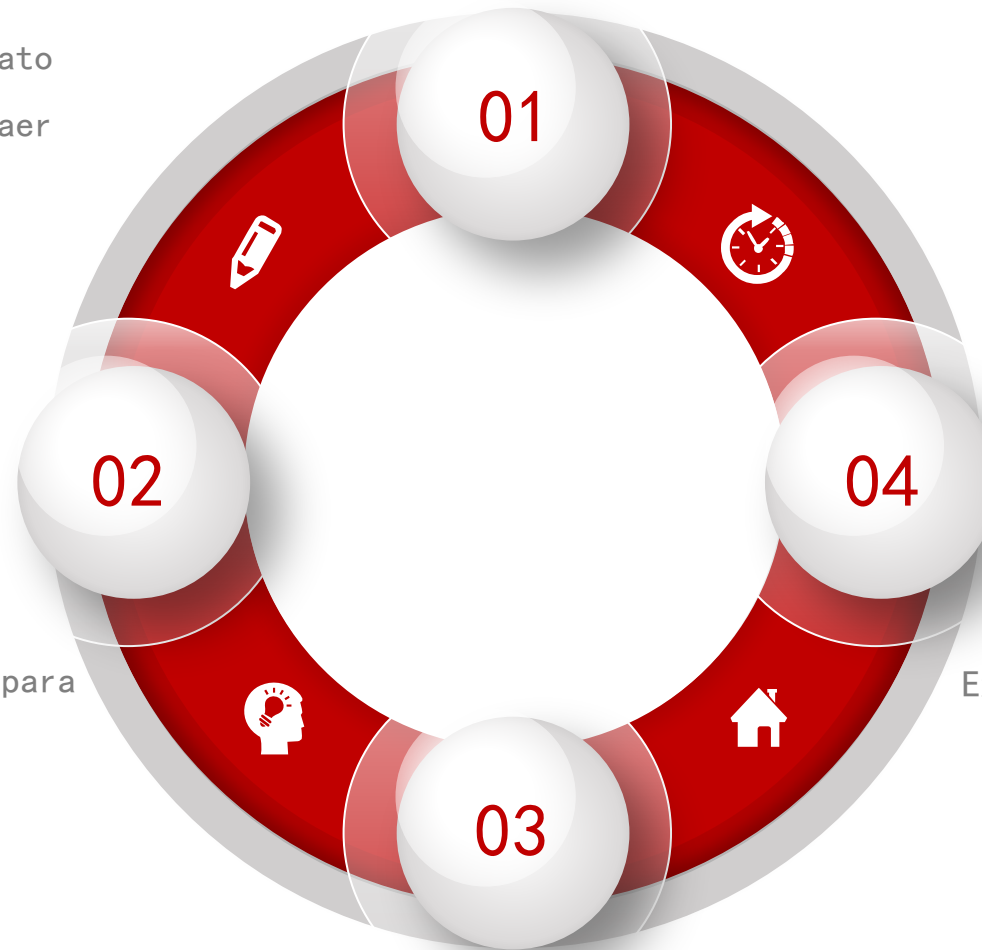
Normalización y extracción de la hora para crear la variable numérica hour\_occ.

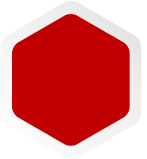
## Selección de Fecha Base

Se determina qué columna de fecha será la referencia para el análisis temporal.

## Normalización de Variables Temporales

Extracción de variables como año, mes, día de la semana y fecha del mes.





# Limpieza de Datos

LOGO



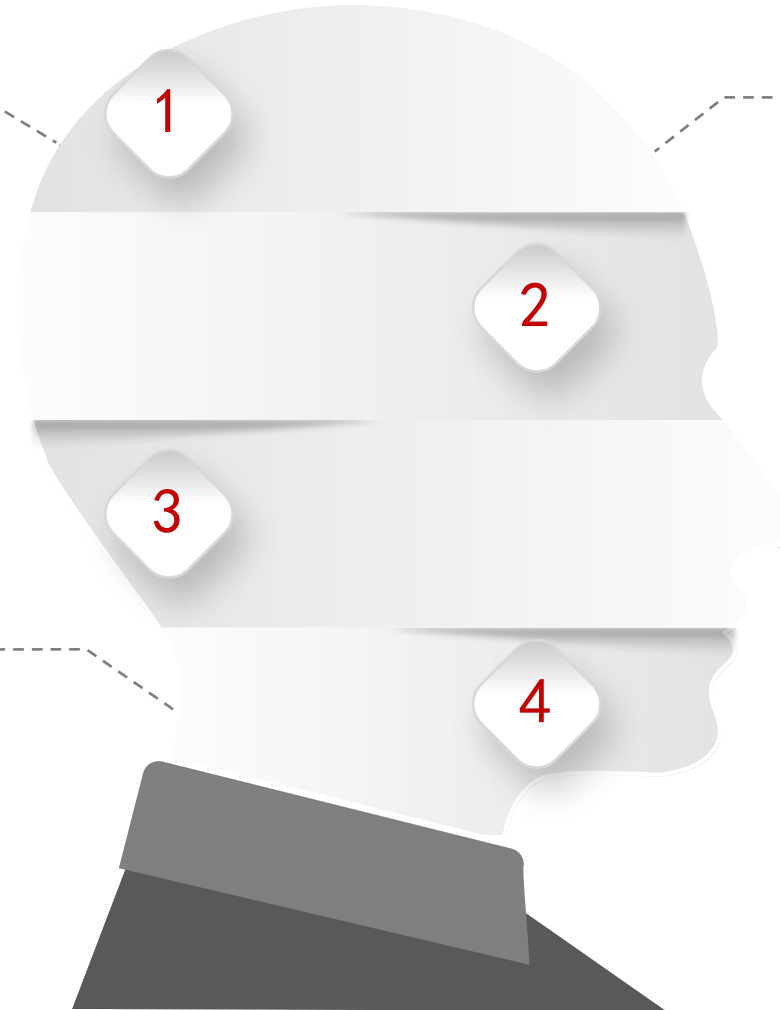
## Limpieza de las coordenadas

Se realiza búsqueda de limpieza de las columnas de latitud y longitud creando un indicador.



## Creación de variable objetivo

Se crea un variable binaria indicando si el delito fue violento para la clasificación.



## Limpieza de Variables Categóricas



Se normalizaron columnas como `area_name` para reducir inconsistencias y variabilidad.



## Limpieza final de registros

Es necesario eliminar registros duplicados y aquellos con falta de datos importantes.



LOGO

# Métodos Utilizados

04

# Análisis Descriptivo

## Correlaciones Numéricas

---

Se realizó un mapa de calor para examinar las correlaciones entre variables numéricas.

## Histograma de Edad de Víctima

---

Se estudió la distribución de la edad de las víctimas.

## Distribución por Día de la Semana

---

Se determinó cómo se distribuyen los incidentes según el día de la semana.

## Tendencia Mensual de Incidentes

---

Se analizó la evolución del número de incidentes por mes.

## Top 15 Tipos de Crimen

---

Se obtuvieron y visualizaron los 15 tipos de delito más frecuentes.

# Modelado Estadístico

Se ajustó un modelo de regresión lineal para explicar las variaciones mensuales.

**Regresión Lineal OLS sobre Incidencias Mensuales**

01

Se graficaron los residuos a lo largo del tiempo.

**Residuos del Modelo OLS**

02

03

Se entrenó un modelo Random Forest con pipeline de procesamiento para clasificación.

**Machine Learning: Random Forest**

05

04

**Ajuste del Modelo OLS**

Se evaluó el ajuste del modelo con un gráfico observado vs predicho.

**Regresión Logística para Odds Ratios**

Se utilizó regresión logística para modelar la probabilidad de que un delito sea violento.

LOGO

# Resultados del Análisis

A large red hexagon with a white border and a subtle drop shadow, containing the white number '05'. It is positioned on the right side of a light gray horizontal band. Two smaller red hexagons are also present: one above and to the right of the main hexagon, and another below and to the left of it.

05



Tipos de crímenes más comunes

**Target**



LOGO

# Conclusiones y Recomendaciones

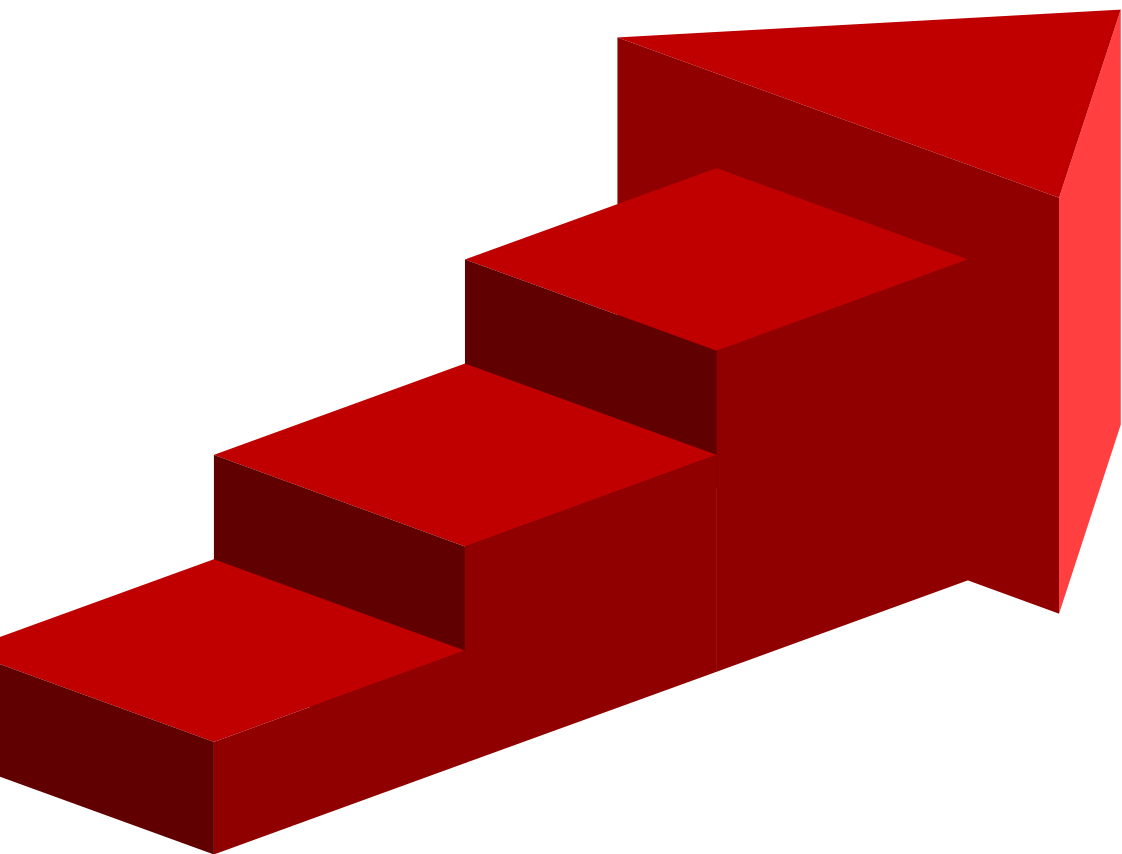
06



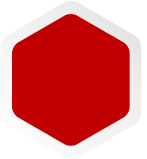
# Conclusiones

---

LOGO



resumen del análisis y resultados obtenidos



Acciones recomendadas basadas en el  
análisis

Target







**SCHOOL OF ENGINEERING**  
UNIDAD DE POSTGRADO FICCT - UAGRM  
PLATAFORMA EDUCATIVA

# Thank you!

