

# ESTUDIO SOBRE EL COMPORTAMIENTO CRIMINAL EN LA CIUDAD DE LOS ÁNGELES - CALIFORNIA DESDE 2020 HASTA 2024

Isaac Amín Sofán hernández<sup>1</sup>, Yan Carlos Tano Espitia<sup>2</sup>, Sergio Andrés Tascon Durango<sup>3</sup>

Facultad de Ingeniería  
Politécnico Jaime Isaza Cadavid  
Apartado, Antioquia, Colombia  
{Isaac\_sofan82192, yan\_tano82192, sergio\_tascon82201}@elpoli.edu.co

## RESUMEN

El estudio analiza el comportamiento criminal en la ciudad de Los Ángeles, California, desde 2020 hasta 2024. Los objetivos principales incluyen la caracterización de las víctimas, la identificación de zonas calientes y la determinación de las temporadas más peligrosas. Utilizando la metodología CRISP-DM para la extracción y análisis de datos de los informes del Departamento de Policía de Los Ángeles (LAPD), se identificaron patrones delictivos. Los hallazgos revelan una mayor incidencia de crímenes en los primeros trimestres de cada año, con armas de fuego y cuchillos siendo las más comunes. Las conclusiones destacan la necesidad de políticas públicas basadas en evidencia para mejorar la seguridad ciudadana.

**Palabras clave:** Comportamiento Criminal; Los Ángeles, Zonas Calientes, Temporadas Peligrosas, Metodología CRISP-DM

## STUDY ON CRIMINAL BEHAVIOR IN THE CITY OF LOS ANGELES - CALIFORNIA FROM 2020 TO 2024

## ABSTRACT

*The study analyzes criminal behavior in the city of Los Angeles, California, from 2020 to 2024. The main objectives include the characterization of victims, the identification of hot spots and the determination of the most dangerous seasons. Using the CRISP-DM methodology for data extraction and analysis from Los Angeles Police Department (LAPD) reports, crime patterns were identified. The findings reveal a higher incidence of crime in the first quarters of each year, with firearms and knives being the most common. The conclusions highlight the need for evidence-based public policies to improve citizen security.*

**Keywords:** Criminal Behavior; Los Angeles, Hot Zones, Dangerous Seasons, CRISP-DM Methodology.



## 1. INTRODUCCIÓN

El comportamiento criminal en las grandes urbes es un tema de creciente interés y preocupación, especialmente en ciudades tan diversas y extensas como Los Ángeles, California. Este estudio se centra en analizar los patrones de criminalidad en Los Ángeles desde 2020 hasta 2024, utilizando la metodología CRISP-DM para la extracción y análisis de datos. A través de la comprensión de los incidentes delictivos, este trabajo busca proporcionar información valiosa para la toma de decisiones estratégicas que mejoren la seguridad pública.

El análisis de la criminalidad en Los Ángeles no solo se justifica por la necesidad de entender los factores que impulsan estos comportamientos, sino también por el impacto que tiene en la calidad de vida de sus habitantes. Los datos utilizados provienen de los informes del Departamento de Policía de Los Ángeles (LAPD), lo que permite un análisis detallado y riguroso de las tendencias y patrones delictivos. Este enfoque ayuda a identificar áreas problemáticas y a sugerir acciones preventivas y correctivas basadas en evidencia.

Este documento aborda la distribución geográfica de los crímenes, las características de las víctimas, así como las armas utilizadas y los lugares más frecuentes de los delitos. A través de técnicas avanzadas de análisis de datos y visualización, se busca proporcionar un panorama claro y comprensible del comportamiento criminal en la ciudad, facilitando así la formulación de políticas públicas más efectivas y la optimización de recursos policiales.

En las siguientes secciones, se detallarán los materiales y métodos utilizados, los resultados obtenidos y se discutirá la relevancia de estos hallazgos en el contexto de la criminología urbana y la gestión de la seguridad pública. Este estudio contribuye a la literatura existente y ofrece una base sólida para futuras investigaciones en el campo de la criminología y la ciencia de datos aplicadas a la seguridad ciudadana.

## 2. MATERIALES Y METODO

En la actualidad, existen múltiples metodologías que facilitan el análisis de datos y la extracción de información relevante. Una de las más reconocidas es CRISP-DM (Cross Industry Standard Process for Data Mining). A pesar de ser ampliamente utilizada en proyectos de minería de datos durante más de veinte años, no goza de la misma notoriedad en muchos entornos laborales, como lo dice [1].

Su metodología de proceso se compone de seis etapas: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación e Implementación o despliegue. La secuencia de estas etapas no es estricta. Cada una se divide en diversas tareas generales de segundo nivel. CRISP-DM define un conjunto de tareas y actividades para cada fase del proyecto, pero no detalla cómo deben ser ejecutadas, tal como nos indica [2]. Además, la Figura 1, nos muestra el flujo iterativo de la metodología.

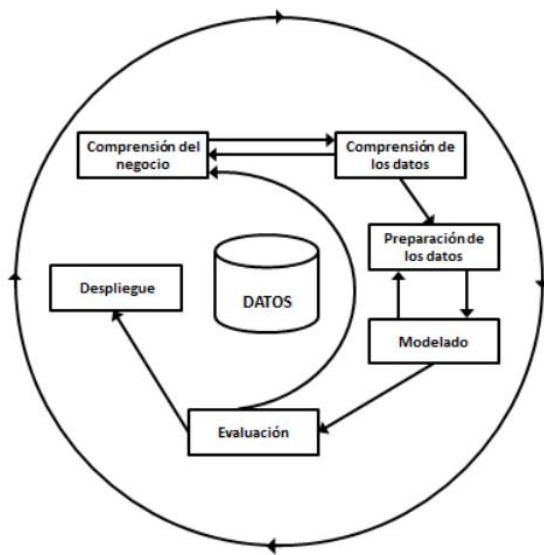


Figura 1. Fases Metodología CRISP DM

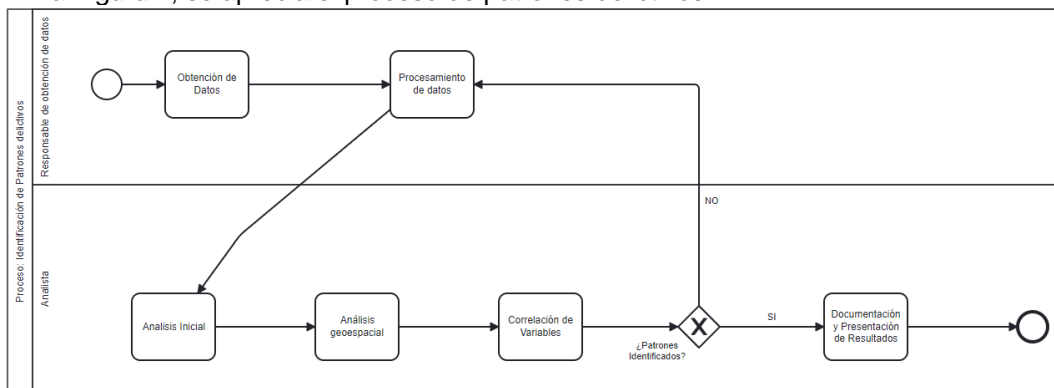
## 2.1. Comprensión del negocio

Se hizo uso de una base de datos proporcionada por el departamento de policía de los Ángeles (LAPD). Es el cuerpo de policía local de la ciudad de Los Ángeles, estado de California, Estados Unidos. Cuenta con alrededor de 9000 policías, cubriendo un área de 470 kilómetros ofrece protección y seguridad a aproximadamente 4 millones de personas.

El objetivo es mejorar la seguridad y la toma de decisiones estratégicas en el departamento de Policía de Los Ángeles mediante la aplicación de la metodología CRIPS DM a los datos de incidentes delictivos en la ciudad desde 2020. Este procedimiento ayudará a identificar patrones, analizar riesgos y sugerir acciones preventivas y correctivas, lo que favorecerá la mejora constante de la seguridad organizacional y la eficiencia operativa.

Con la finalidad de identificar las variables clave en los datos de incidentes delictivos y analizar la distribución geográfica de los sucesos para detectar patrones y tendencias.

En la *Figura 2*, se aprecia el proceso de patrones delictivos.



*Figura 2. Diagrama BPMN (Business Process Model and Notation)*

## 2.2. Comprensión de los datos

El conjunto de datos incluye incidentes criminales en la Ciudad de Los Ángeles, Estados Unidos, desde 2020 hasta ahora. La actualización más reciente fue en marzo de 2024 y es bimestral. Estos datos se extraen de informes de delitos originales en papel. El Departamento de Policía de Los Ángeles proporciona este conjunto de datos. LAPD OpenData es el dueño del conjunto de datos. Los encabezados del DataSet se muestran mediante una exploración inicial realizada en Python, como se muestra en la *Figura 3*.

| ID | RegisterID | Date      | Time       | AreaID | AreaName | CrimeID   | CrimeName | CrimeSceneID                               | CrimeSceneName | WeaponUsedID                                 | WeaponUsedName | VictimAge | VictimSex | VictimDescent |   |
|----|------------|-----------|------------|--------|----------|-----------|-----------|--|----------------|--|----------------|-----------|-----------|---------------|---|
| 0  | 1          | 190326475 | 03/01/2020 | 2130   | 7        | Wilshire  | 510       | VEHICLE - STOLEN                           | 101.0          | STREET                                       | 0              | NaN       | 0         | M             | O |
| 1  | 2          | 200106753 | 02/08/2020 | 1800   | 1        | Central   | 330       | BURGLARY FROM VEHICLE                      | 128.0          | BUS STOP/LAYOVER (ALSO QUERY 124)            | 0              | NaN       | 47        | M             | O |
| 2  | 3          | 200320258 | 11/04/2020 | 1700   | 3        | Southwest | 480       | BIKE - STOLEN                              | 502.0          | MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC) | 0              | NaN       | 19        | X             | X |
| 3  | 4          | 200907217 | 03/10/2020 | 2037   | 9        | Van Nuys  | 343       | SHOPLIFTING- GRAND THEFT (\$950.01 & OVER) | 405.0          | CLOTHING STORE                               | 0              | NaN       | 19        | M             | O |
| 4  | 5          | 220614831 | 05/08/2021 | 1200   | 6        | Hollywood | 354       | THEFT OF IDENTITY                          | 102.0          | SIDEWALK                                     | 0              | NaN       | 28        | M             | H |

*Figura 3. Encabezados de DataSet Crimes*

Al realizar una exploración más precisa, empleando la librería Pandas de Python, obtuvimos información más detallada del DataSet, peso del archivo, columnas, total de registros [filas] datos y tipos de datos (categóricos, enteros, flotantes), esta información se expone en la *Figura 4*.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 883987 entries, 0 to 883986
Data columns (total 15 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   ID               883987 non-null  int64
1   RegisterID       883987 non-null  int64
2   Date            883987 non-null  object
3   Time            883987 non-null  int64
4   AreaID           883987 non-null  int64
5   AreaName         883987 non-null  object
6   CrimeID          883987 non-null  int64
7   CrimeName        883987 non-null  object
8   CrimeSceneID     883977 non-null  float64
9   CrimeSceneName   883447 non-null  object
10  WeaponUsedID     883987 non-null  int64
11  WeaponUsedName   306604 non-null  object
12  VictimAge        883987 non-null  int64
13  VictimSex        883987 non-null  object
14  VictimDescent    883987 non-null  object
dtypes: float64(1), int64(7), object(7)
memory usage: 101.2+ MB
```

Figura 4. Información Inicial DataSet Crimes

Además, al realizar un análisis estadístico básico y presentarlo en forma de matriz transpuesta, tal como se presenta en la Figura 5.

|              | count    | mean         | std          | min   | 25%         | 50%         | 75%         | max         |
|--------------|----------|--------------|--------------|-------|-------------|-------------|-------------|-------------|
| ID           | 883987.0 | 4.419940e+05 | 2.551852e+05 | 1.0   | 220997.5    | 441994.0    | 662990.5    | 883987.0    |
| RegisterID   | 883987.0 | 2.173818e+08 | 1.139578e+07 | 817.0 | 210314548.5 | 220322187.0 | 230221637.5 | 249904551.0 |
| Time         | 883987.0 | 1.336561e+03 | 6.532328e+02 | 1.0   | 900.0       | 1415.0      | 1900.0      | 2359.0      |
| AreaID       | 883987.0 | 1.070311e+01 | 6.100755e+00 | 1.0   | 6.0         | 11.0        | 16.0        | 21.0        |
| CrimeID      | 883987.0 | 5.007581e+02 | 2.076298e+02 | 110.0 | 331.0       | 442.0       | 626.0       | 956.0       |
| CrimeSceneID | 883977.0 | 3.062995e+02 | 2.171741e+02 | 101.0 | 101.0       | 203.0       | 501.0       | 976.0       |
| WeaponUsedID | 883987.0 | 1.260156e+02 | 1.876490e+02 | 0.0   | 0.0         | 0.0         | 400.0       | 516.0       |
| VictimAge    | 883987.0 | 2.966063e+01 | 2.182583e+01 | -3.0  | 0.0         | 31.0        | 45.0        | 120.0       |

Figura 5. Análisis Estadístico Básico del DataSet Crimes

Consecuentemente, hemos examinado el sexo de las víctimas en relación al porcentaje sobre las incidencias de los crímenes. Igualmente se formalizó un diagnostico descriptivo sobre los datos de categoría 'object', donde se aprecian el recuento total, así mismo el total de datos únicos, su frecuencia y los resultados se presentan en la Figura 6.

|                | count  | unique | top  | freq   |
|----------------|--------|--------|--|--------|
| Date           | 883987 | 800    | 02/02/2023                                     | 2283   |
| AreaName       | 883987 | 21     | Central  | 60123  |
| CrimeName      | 883987 | 138    | VEHICLE - STOLEN                               | 94923  |
| CrimeSceneName | 883447 | 306    | STREET   | 224004 |
| WeaponUsedName | 306604 | 79     | STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE) | 164310 |
| VictimSex      | 883987 | 5      | M  | 363260 |
| VictimDescent  | 883987 | 20     | H  | 269743 |

Figura 6. Análisis Descriptivo Datos Categóricos Tipo object

Posteriormente, se identificaron todos los datos categóricos únicos del tipo 'object'. Este proceso permitió clarificar y categorizar la información disponible, facilitando así un análisis más estructurado y preciso. A través de esta identificación y clasificación, se obtuvo un entendimiento completo de los datos, permitiendo una visión integral y detallada de los distintos elementos presentes en el conjunto de datos. Este entendimiento es fundamental para garantizar la calidad y relevancia del análisis posterior. Los resultados de este proceso se reflejan en la Tabla 1, que presenta el diccionario de datos del dataset, proporcionando una descripción exhaustiva de las variables y sus características.

| Column Name    | Description  | Type        |
|----------------|--|-------------|
| DR_NO          | Número de División de Registros: Número de expediente oficial compuesto por un año de 2 dígitos, identificación de área y 5 dígitos  | Plain Text  |
| Date Rptd      | DD/MM/YYYY   | Date & Time |
| DATE OCC       | DD/MM/YYYY   | Date & Time |
| TIME OCC       | En horario militar de 24 horas.  | Plain Text  |
| AREA           | El Departamento de Policía de Los Ángeles (LAPD, por sus siglas en inglés) tiene 21 Estaciones de Policía Comunitarias denominadas Áreas Geográficas dentro del departamento. Estas áreas geográficas están numeradas secuencialmente del 1 al 21.   | Plain Text  |
| AREA NAME      | Las 21 Áreas Geográficas o Divisiones de Patrulla también reciben una designación de nombre que hace referencia a un punto de referencia o a la comunidad circundante de la que es responsable. Por ejemplo, la División de la Calle 77 se encuentra en la intersección de South Broadway y la Calle 77, sirviendo a los vecindarios del sur de Los Ángeles.   | Plain Text  |
| Rpt Dist No    | Un código de cuatro dígitos que representa una subárea dentro de un área geográfica. Todos los registros de delitos hacen referencia al "RD" en el que ocurrió para las comparaciones estadísticas. Encuentre los distritos de informes de LAPD en el GeoHub de la ciudad de Los Ángeles en <a href="http://geohub.lacity.org/datasets/c4f83909b81d4786aa8ba8a74a4b4db1_4">http://geohub.lacity.org/datasets/c4f83909b81d4786aa8ba8a74a4b4db1_4</a>                                      | Plain Text  |
| Part 1-2       | Parte de la Ciudad   | Number      |
| Crm Cd         | Indica el delito cometido. (Igual que el Código Penal 1)   | Plain Text  |
| Crm Cd Desc    | Define el Código Penal previsto.   | Plain Text  |
| Mocodes        | Modus Operandi: Actividades asociadas con el sospechoso en la comisión del delito. Consulte el PDF adjunto para obtener una lista de códigos MO en orden numérico. <a href="https://data.lacity.org/api/views/y8tr-7khq/files/3a967fbd-f210-4857-bc52-60230efe256c?download=true&amp;filename=MO%20CODES%20(número%20orden).pdf">https://data.lacity.org/api/views/y8tr-7khq/files/3a967fbd-f210-4857-bc52-60230efe256c?download=true&amp;filename=MO%20CODES%20(número%20orden).pdf</a> | Plain Text  |
| Vict Age       | Dos caracteres numéricos   | Plain Text  |
| Vict Sex       | F - Female M - Male X - Unknown  | Plain Text  |
| Vict Descent   | Código de descendencia: A - Otro Asiático B - Negro C - Chino D - Camboyano F - Filipino G - Guamaní H - Hispano/Latino/Mexicano I - Indio Americano/Nativo de Alaska J - Japonés K - Coreano L - Laosiano O - Otro P - Isleño del Pacífico S - Samoano U - Hawaiano V - Vietnamita W - Blanco X - Desconocido Z - Asiático Indio  | Plain Text  |
| Premis Cd      | El tipo de estructura, vehículo o lugar donde ocurrió el delito.   | Number      |
| Premis Desc    | Define el código de premisa proporcionado.   | Plain Text  |
| Weapon Used Cd | Define el código de premisa proporcionado.   | Plain Text  |
| Weapon Desc    | Define el código de arma utilizada proporcionado.  | Plain Text  |
| Status         | Estado del caso. (IC es el valor predeterminado)   | Plain Text  |
| Status Desc    | Define el código de estado proporcionado.  | Plain Text  |
| Crm Cd 1       | Indica el delito cometido. El Código Penal 1 es el principal y el más grave. Los Códigos Penales 2, 3 y 4 son, respectivamente, delitos menos graves. Las cifras más bajas de la clase de delincuencia son más graves.   | Plain Text  |
| Crm Cd 2       | Puede contener un código para un delito adicional, menos grave que el Código Penal 1.  | Plain Text  |
| Crm Cd 3       | Puede contener un código para un delito adicional, menos grave que el Código Penal 1.  | Plain Text  |
| Crm Cd 4       | Puede contener un código para un delito adicional, menos grave que el Código Penal 1.  | Plain Text  |
| LOCATION       | La dirección de la calle del incidente del delito se redondeó a la cuadra cien más cercana para mantener el anonimato.   | Plain Text  |
| Cross Street   | Cruce de la calle de la dirección redondeada   | Plain Text  |
| LAT            | Latitude   | Number      |
| LON            | Longitude  | Number      |

Tabla 1. Diccionario de Datos del DataSet Crimes

### 2.3. Preparación de los datos

Del paso anterior tenemos los siguientes Insights, los cuales son nuestro punto de partida para realizar la limpieza correspondiente de los datos:

- Ajustar datos con valores nulos en las columnas: 'CrimeSceneId' rellenar con 0, 'CrimeSceneName' y 'WeaponUsedName' rellenar con 'NO APLICA'.
- Convertir el tipo de dato de la columna 'CrimeSceneID' de float a int.
- Ajustar datos atípicos en la columna 'VictimAge', la edad de la víctima tiene edades menores a 0, es decir, edades negativas y tal cosa no puede existir.
- Convertir los datos de la columna 'AreaName' que están en minúsculas a mayúsculas.
- Ajustar datos atípicos en la columna 'Time' donde encontramos según el formato usado en el DataSet horas menores a 100 "1:00 am" según formato militar. Datos como 1,5, 20, 45 se multiplican por un factor para estandarizar las horas a partir de 100 y evitar valores menores.
- La columna 'VictimSex' es el sexo de la víctima, en un principio se habla de que 'M-Masculino', 'F-Femenino' y 'X-Desconocido' pero al explorar el DataSet encontramos otros datos como 'H' y '-' de los cuales se desconoce su significado por lo que se asume a partir del momento que los registros con datos con estos últimos valores se deben reemplazar con 'X-Desconocido'.
- La columna 'VictimDescent' contiene los valores sobre la descendencia de las víctimas y usa un valor 'X-Unknown' para datos desconocidos, pero se encontró un valor '-' que no tiene significado por lo cual esos datos se estandarizan pasando a tomar el valor de 'X'.
- Los valores de la columna 'VictimSex' se reemplazan por su significado, 'M' por 'MALE', 'F' por 'FEMALE' y 'X' por 'UNKNOWN'. Esto es para que la información sea más clara y precisa.

- Los valores de la columna 'VictimDescent' se reemplazan por su significado, 'A' por 'OTHER ASIAN', 'B' por 'BLACK', 'W' por 'WHITE'. Esto es para que la información sea más clara y precisa.
- Realizar correlaciones para rellenar registros faltantes, nulos y atípicos.

Al realizar el proceso de exploración y análisis del DataSet se hace necesario eliminar algunas columnas y reemplazar sus nombres en vista de que no son relevantes y se empezó con la transformación de los datos:

- Se encontraron fechas con formatos diferentes por lo que se procedió a la transformación de las fechas del formato "mm/dd/yyyy" al formato "dd/mm/yyyy".
- Se encontró la hora en un formato que dificulta su entendimiento por lo que es necesario transformarlo a hora en formato militar para mejor comprensión de los datos.
- Se ajustaron los datos con valores nulos en las columnas: 'CrimeSceneId' rellenar con 0, 'CrimeSceneName' y 'WeaponUsedName' rellenar con 'NO APLICA'.
- Se Ajustaron los datos atípicos en la columna 'VictimAge', la edad de la víctima tiene edades menores a 0, es decir, edades negativas y tal cosa no puede existir.
- Se transformaron los datos de la columna 'AreaName' que están en minúsculas a mayúsculas.
- Se ajustaron datos atípicos en la columna 'Time' donde encontramos según el formato usado en el DataSet horas menores a 100 "1:00 am" según formato militar. Datos como 1,5, 20, 45 se multiplican por un factor para estandarizar las horas a partir de 100 y evitar valores menores.
- Para la columna 'VictimSex' que es el sexo de la víctima, en un principio se habla de que 'M-Masculino', 'F-Femenino' y 'X-Desconocido' pero al explorar el DataSet encontramos otros datos como 'H' y '-' de los cuales se desconoce su significado por lo que se asume a partir del momento que los registros con datos con estos últimos valores se deben reemplazar con 'X-Desconocido'.
- Para la columna 'VictimDescent' la cual contiene los valores sobre la descendencia de las víctimas y usa un valor 'X-Unknown' para datos desconocidos, pero se encontró un valor '-' que no tiene significado por lo cual esos datos se estandarizan pasando a tomar el valor de 'X'.
- Se realizaron las correlaciones para rellenar registros faltantes, nulos y atípicos.

Se aplica la transformación de manera uniforme en todo el conjunto de datos.

En el manejo de datos nulos y atípicos, Se evaluó la naturaleza y la importancia de los campos con datos nulos. Se decide si eliminar los campos con datos nulos, teniendo en cuenta los valores faltantes o dejar estos campos con datos nulos según lo que sea necesario.

Luego de realizar el proceso de exploración y análisis del DataSet se procedió a eliminar algunas columnas y reemplazar sus nombres en vista de que no son de relevancia para alcanzar los objetivos planteados

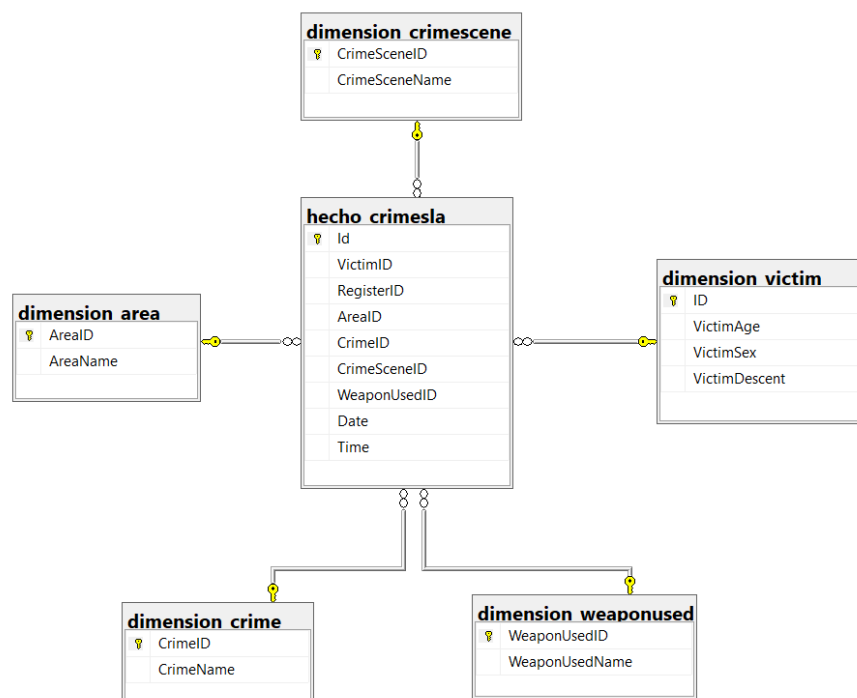
## 2.4. Modelado

En la fase de modelado, se destaca el modelo dimensional o de estrella. El diseño del almacén de datos dimensional se basa en el esquema en estrella. El término 'esquema en estrella' hace referencia a la combinación de una tabla de hechos con varias tablas dimensionales.

En nuestro esquema tenemos una tabla hecho, en la cual se busca Identificar las áreas geográficas, crímenes, escenas del crimen y armas empleadas en la comisión de los crímenes en la ciudad de Los Ángeles, así como también otras características propias de las víctimas como edad, sexo, descendencia desde el año 2020 hasta el presente. Además, de cinco tablas dimensiones:

1. **dimension\_area:** Corresponde a la dimensión sobre el área o división geográfica de la ciudad de Los Ángeles realizada por el departamento de policía (LADP), para identificar en que parte de la ciudad ocurrió el crimen y nos servirá para conocer las áreas con índices de criminalidad más alto y bajo.
2. **dimensión\_crime:** Corresponde a la dimensión sobre el crimen, donde se describen los crímenes cometidos como robo, secuestro, homicidio, entre otros, que nos servirá para identificar que crimines son más frecuentes o no.
3. **dimension\_crimescene:** Corresponde a la dimensión sobre la escena del crimen, se especifica el lugar donde ocurrieron los hechos como en un carro, casa, calle, parque, entre otros. Nos servirá para identificar la estructura o el lugar donde se presentan los crímenes.
4. **dimension\_weaponused:** Corresponde a la dimensión sobre el arma usada, se describe el arma empleada para cometer el crimen como pistola 9m, fusil, cuchillo, manos, entre otros. Nos servirá para identificar que elementos son más utilizados para cometer los crímenes.
5. **dimension\_victim:** Corresponde a la dimensión sobre la víctima donde idvictima, edad y características propias de la víctima como el sexo y la descendencia (grupo poblacional usado

en Estados Unidos; blanco, latino, asiático, entre otros) que nos aportan una idea sobre la persona que padeció el crimen. Nuestro modelo dimensional final se detalla a profundidad en la *Figura 7*.



*Figura 7. Modelo Dimensional Crimes*

## 2.5. Evaluación

El modelo dimensional está bien diseñado y es efectivo para almacenar y analizar datos sobre crímenes en la ciudad de Los Ángeles. El modelo tiene las siguientes características positivas:

- Las tablas de dimensiones están bien normalizadas. Cada tabla almacena un conjunto de datos distinto y coherente, y las tablas no están duplicadas.
- Las tablas de dimensiones están relacionadas entre sí de forma lógica. Las tablas se relacionan entre sí mediante claves foráneas, lo que permite consultar y analizar los datos de forma eficaz.
- El modelo es flexible y ampliable. Se pueden añadir nuevas tablas de dimensiones al modelo a medida que se necesiten datos adicionales.

En consecuencia, el modelo dimensional es una herramienta eficaz para almacenar y analizar datos sobre crímenes en la ciudad de Los Ángeles. El modelo tiene un diseño sólido y es flexible y ampliable.

## 2.6. Implementación

Se realizó la extracción, transformación y carga (ETL) de datos desde el origen (archivo “crimes.csv”) hasta su destino el sistema de bases de datos Microsoft SQL Server mediante la importación de datos al convertir el archivo ‘csv’ a ‘xlsx’ para facilitar el proceso.

1. Se creó en Microsoft SQL Server una base de datos llamada ‘DBIncidencias\_crímenes\_LA’, donde se crearon las 5 tablas (dimensiones Area, Crime, CrimeScene, Victim, WeaponUsed) y la tabla de hechos (hecho\_crimsesla).
2. Para realizar la importación de los datos para las 5 dimensiones, se hace uso del asistente de Microsoft SQL Server para la importación de datos desde un archivo de ‘csv’, el cual se convierte en una tabla dentro de la base de datos llama ‘crimes’.
3. Luego, se crea un procedimiento de almacenado (un programa que se ejecuta dentro de la base de datos) se realiza la migración de los datos desde la tabla ‘crimes’ que contiene todos los datos que estaban en el archivo ‘csv’ a cada una de las tablas (dimensiones) creadas y a la tabla de hechos ingresando solamente en las dimensiones los datos distintos que alimentan la tabla de hechos.
4. Se ejecuta el procedimiento almacenado “migracionDatos” mediante el comando “exec migracionDatos” y comienza el proceso de carga de datos a las distintas tablas.



El procedimiento de almacenado se presenta en la *Figura 8*.

```
CREATE PROCEDURE migracionDatos
@fechaInicio date, @fechaFin date
AS
-- Insertar Información en la dimensión Area
INSERT INTO dimension_area (AreaID,AreaName)
SELECT distinct crimes.AreaID, crimes.AreaName
FROM crimes LEFT OUTER JOIN dimension_area ON dimension_area.AreaID=crimes.AreaID
WHERE dimension_area.AreaID IS NULL

-- Insertar Información en la dimensión Crime
INSERT INTO dimension_crime (CrimeID,CrimeName)
SELECT distinct crimes.CrimeID, crimes.CrimeName
FROM crimes LEFT OUTER JOIN dimension_crime ON dimension_crime.CrimeID=crimes.CrimeID
WHERE dimension_crime.CrimeID IS NULL

-- Insertar Información en la dimensión CrimeScene
INSERT INTO dimension_crimescene (CrimeSceneID,CrimeSceneName)
SELECT distinct crimes.CrimeSceneID,crimes.CrimeSceneName
FROM crimes left OUTER JOIN dimension_crimescene ON dimension_crimescene.CrimeSceneID=crimes.CrimeSceneID
WHERE dimension_crimescene.CrimeSceneID IS NULL

-- Insertar Información en la dimensión WeaponUsed
INSERT INTO dimension_weaponused (WeaponUsedID,WeaponUsedName)
SELECT distinct crimes.WeaponUsedID, crimes.WeaponUsedName
FROM crimes LEFT OUTER JOIN dimension_weaponused ON dimension_weaponused.WeaponUsedID=crimes.WeaponUsedID
WHERE dimension_weaponused.WeaponUsedID IS NULL

-- Insertar Información en la dimensión Victim
INSERT INTO dimension_victim (ID,VictimAge,VictimSex,VictimDescent)
SELECT distinct crimes.ID, crimes.VictimAge, crimes.VictimSex, crimes.VictimDescent
FROM crimes LEFT OUTER JOIN dimension_victim ON dimension_victim.ID=crimes.ID
WHERE dimension_victim.ID IS NULL

-- Insertar Información en la tabla hecho_crimesla
INSERT INTO hecho_crimesla
(VictimID,RegisterID, AreaID, CrimeID, CrimeSceneID, WeaponUsedID, Date, Time)
SELECT crimes.ID, crimes.RegisterID, crimes.AreaID, crimes.CrimeID, crimes.CrimeSceneID, crimes.WeaponUsedID, crimes.Date, crimes.Time
FROM crimes
WHERE crimes.Date>=@fechaInicio and crimes.Date<=@fechaFin;

/*Ejecutando el algoritmo de Almacenado*/
EXECUTE migracionDatos '2020-01-01', '2024-03-31';
-- EXECUTE migracionDatos;
```

Figura 8. Procedimiento de Almacenado Para Modelo Dimensional

Consecutivamente, realizamos consultas en lenguaje SQL para verificar los resultados de la inserción de los datos a través del procedimiento almacenado. En la tabla 'hecho\_crimesla' se visualizan los resultados, y la *Figura 9* detalla esto.

SELECT TOP (10) \*  
FROM hecho\_crimesla

0.00 %

ResultsMessages

|    | Id | VictimID | RegisterID | AreaID | CrimeID | CrimeSceneID | WeaponUsedID | Date       | Time |
|----|----|----------|------------|--------|---------|--------------|--------------|------------|------|
| 1  | 1  | 564789   | 231605322  | 16     | 821     | 502          | 400          | 2022-01-01 | 1200 |
| 2  | 2  | 564790   | 221613180  | 16     | 420     | 101          | 0            | 2022-08-10 | 900  |
| 3  | 3  | 564791   | 220707519  | 7      | 331     | 101          | 0            | 2022-02-04 | 1400 |
| 4  | 4  | 564792   | 220415562  | 4      | 520     | 101          | 0            | 2023-11-02 | 1700 |
| 5  | 5  | 564793   | 221607920  | 16     | 510     | 101          | 0            | 2022-03-05 | 2330 |
| 6  | 6  | 564794   | 231719078  | 17     | 860     | 721          | 400          | 2022-01-01 | 1200 |
| 7  | 7  | 564795   | 221415757  | 14     | 420     | 108          | 0            | 2024-08-03 | 100  |
| 8  | 8  | 564796   | 221309708  | 13     | 901     | 501          | 0            | 2023-04-04 | 810  |
| 9  | 9  | 564797   | 220123008  | 1      | 330     | 123          | 0            | 2022-08-10 | 2230 |
| 10 | 10 | 564798   | 231504375  | 15     | 310     | 501          | 0            | 2024-11-05 | 1200 |

Figura 9. Consulta SQL a tabla 'hecho\_crimesla'



### 3. RESULTADOS

Luego de explorar los informes realizados en Power BI, se identificaron los siguientes hallazgos: De acuerdo con la *Figura 10* crímenes por años, teniendo en cuenta desde el 2020 hasta el 2024, podemos observar que, en los años 2021, 2022 y 2023 se presentaron las cifras más altas de incidentes criminales.

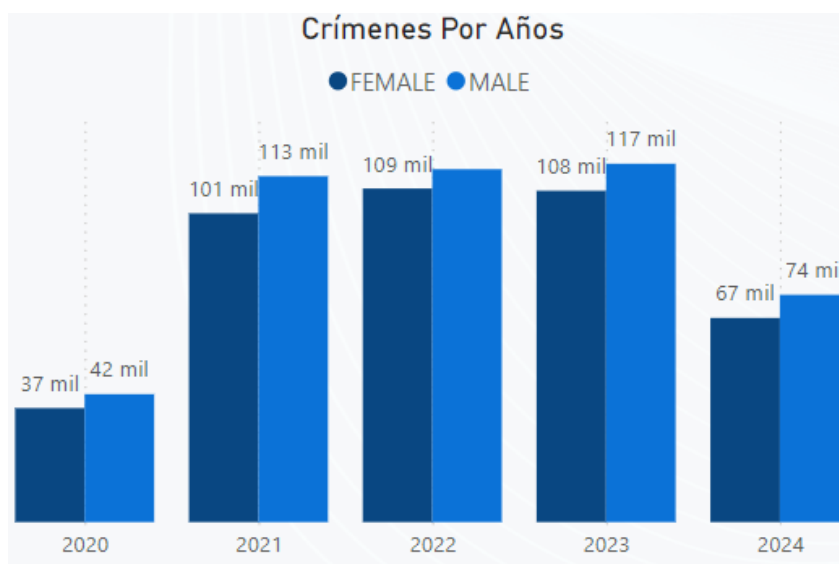


Figura 10. Crímenes por Año

En relación con el trimestre con más incidencias criminales, podemos observar que, en promedio, desde 2020 hasta 2024, en el primer trimestre, comprendido por los meses de enero, febrero y marzo, se presentaron las tasas más altas de crímenes.

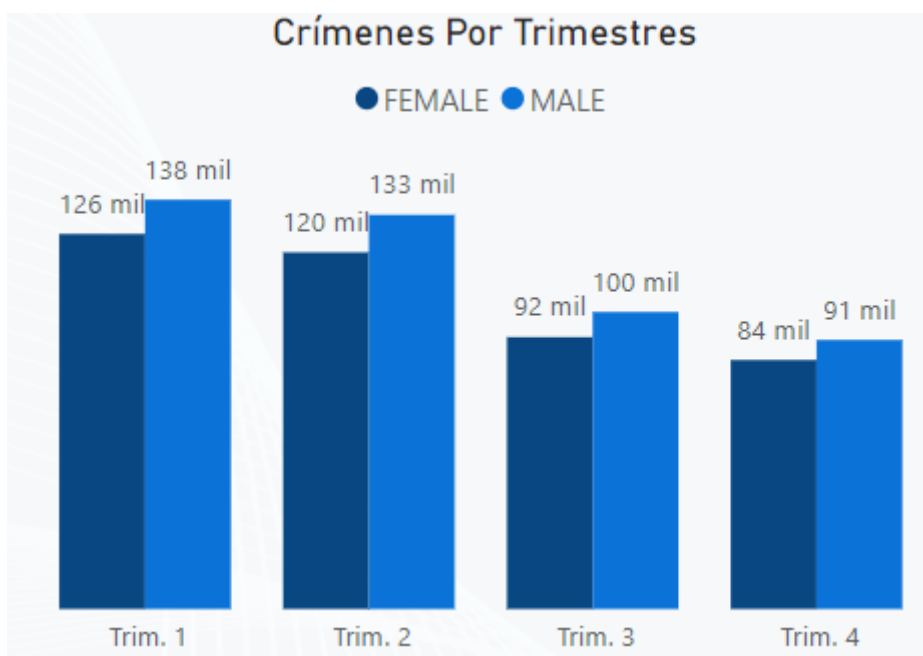


Figura 11. Crímenes por Trimestres

Con respecto a las horas críticas donde la seguridad se ve superada por el número de incidencias, la *Figura 12* evidencia cómo las 12 p.m. es la hora del período comprendido entre el año 2020 y el primer trimestre de 2024 en la que se cometieron más crímenes en la ciudad de Los Ángeles. Esto indica que, a esa hora, en la que generalmente se disponen a almorzar, algunas zonas de la ciudad quedan sin vigilancia, lo que favorece la comisión de delitos. También se presentan muchos casos en horas como las 5 p.m., 6 p.m. y 8 p.m., donde hay un aumento del tráfico de personas dirigiéndose de su lugar de trabajo a casa. Este comportamiento es contrario al ocurrido a las 2 p.m., hora de inicio de la segunda parte de un día laboral, donde se presenta el menor número de delitos.



Figura 12. Top Horas Más Populares

El grupo poblacional al que pertenecen las víctimas es otro factor que se tuvo en cuenta para caracterizarlas. La *Figura 13*, nos permite ver cómo las personas pertenecientes al grupo "HISPANIC" (hispanos) son el grupo más afectado, con alrededor de 280 mil víctimas (25% del total de incidencias) entre 2020 y el primer trimestre de 2024. Luego, encontramos a los "WHITE" (blancos) seguidos de los "BLACK" (afroamericanos). La línea desciende hasta los grupos poblacionales con raíces asiáticas, que son los que menos víctimas registraron según los datos de la policía de Los Ángeles.

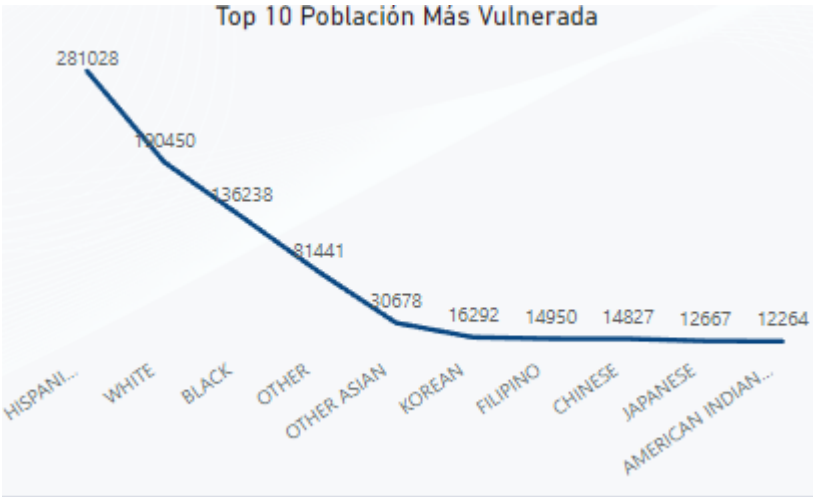


Figura 13. Top 10 Población Más Vulnerada

A continuación, se visualiza la *Figura 14* sobre el top 10 de los crímenes más populares, segmentados por el sexo de la víctima, donde ocupa el primer lugar el "VEHICLE-STOLEN" (robo de vehículo). En una ciudad como Los Ángeles, donde es indispensable un automóvil para poder desplazarse de un lugar a otro debido a las grandes distancias del territorio, encontramos el "THEFT OF IDENTITY" (robo de identidad) en la tercera posición, donde el 59% de las víctimas son mujeres y el 49%, hombres. En la novena posición aparece "INTIMATE PARTNER" (intimidación a la pareja), donde

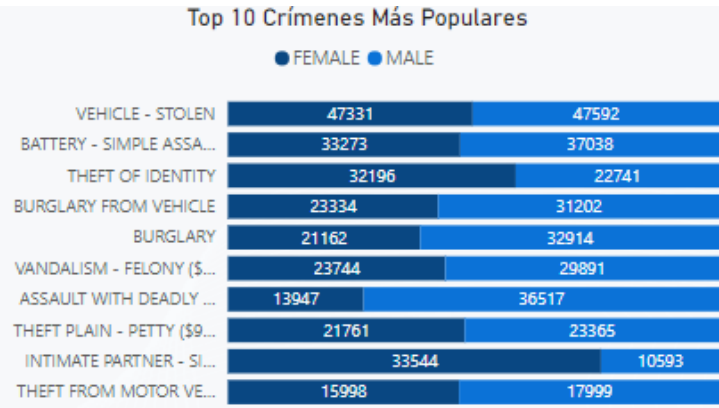
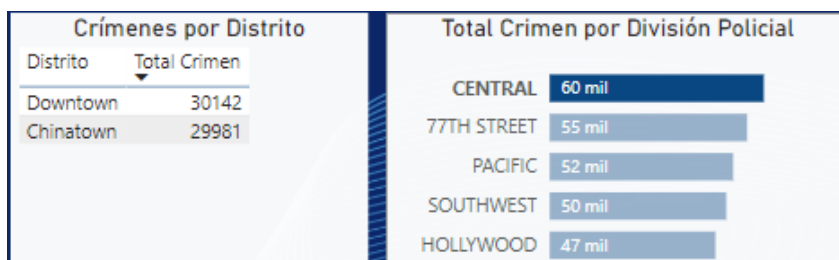


Figura 14. Top Crímenes Más Populares

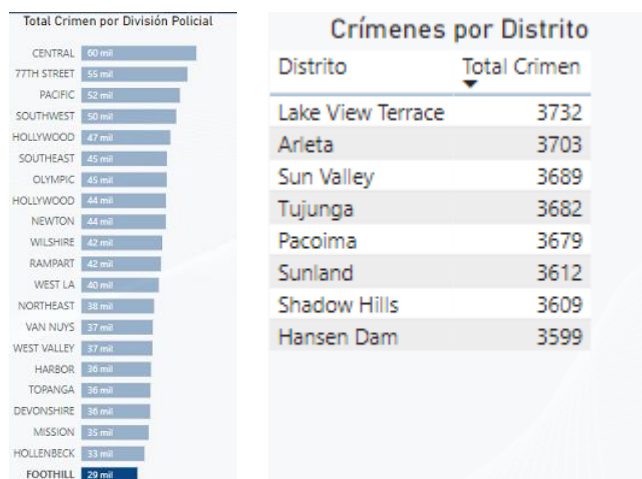
podemos ver un componente de violencia de género, ya que el 76% de las víctimas fueron mujeres y el 24%, hombres.

El área o división policial que más incidencias criminales tuvo a cargo fue "CENTRAL", con 60,123 incidencias, la cual tiene jurisdicción sobre los distritos de la ciudad de Los Ángeles: "Downtown", con 30,142 incidencias, y "Chinatown", con 29,981 incidencias criminales, tal como lo muestra la *Figura 15*.



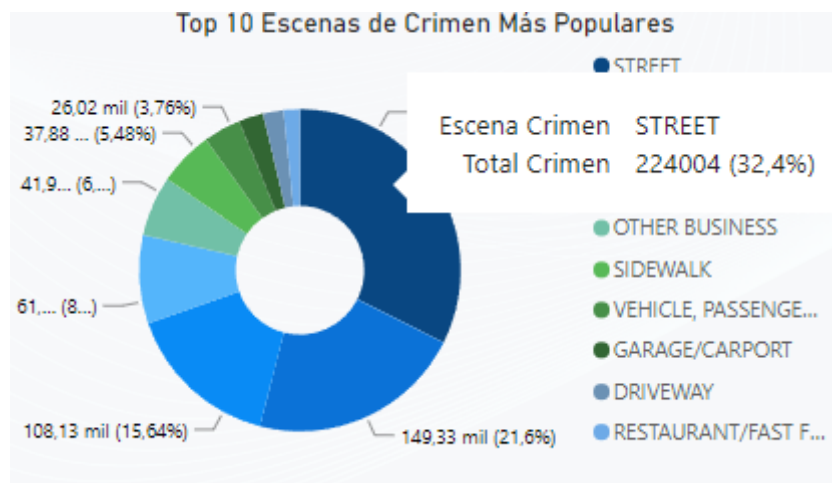
*Figura 15. Crímenes División Policial 'CENTRAL'*

En contraposición, en la *Figura 16*, podemos ver que la estación "FOOTHILL", con jurisdicción sobre 8 distritos, tiene el menor índice con 29,305 incidencias criminales.



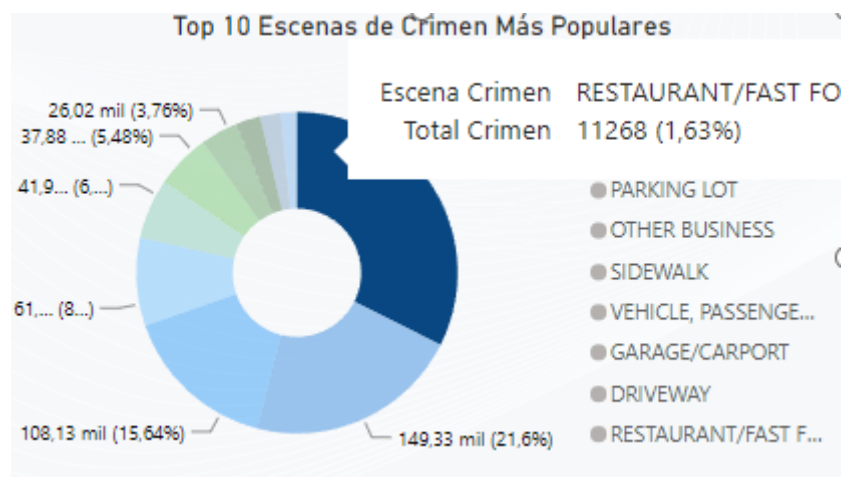
*Figura 16. Crímenes División Policial 'FOOTHILL'*

En la categoría de zonas calientes, encontramos que en el top 10 de las escenas de ocurrencia de crímenes, la más frecuente es "STREET", lo que significa que el 32% de las incidencias criminales ocurren en la calle, como se evidencia en la *Figura 17*.



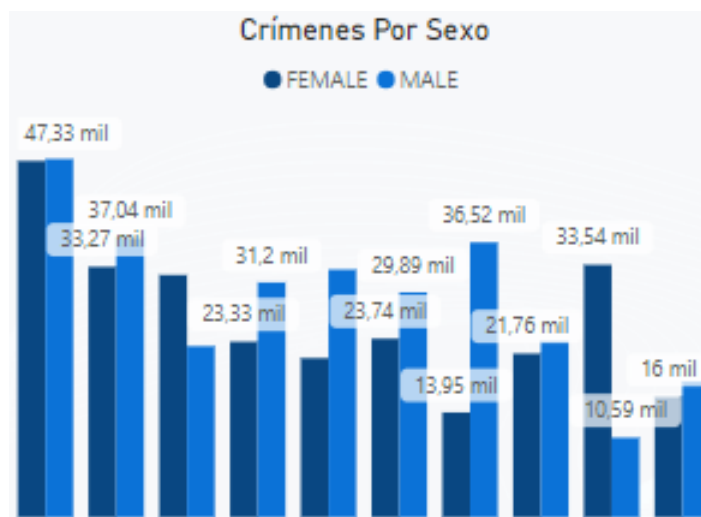
*Figura 17. Escena de Crimen Más Popular*

Contrario al lugar más propicio para cometer un crimen, dentro del top 10, encontramos que en "RESTAURANT/FAST FOOD" (espacios de consumo de comida) es el lugar donde menos incidencias ocurrieron, con apenas el 1.6%, como refleja la *Figura 18*.



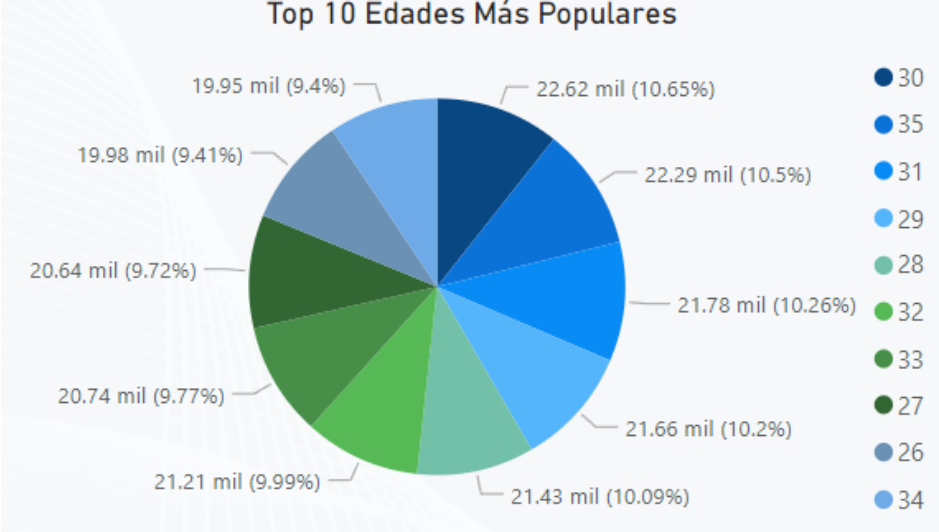
*Figura 18. Escena de Crimen Menos Popular*

La *Figura 19* muestra la cantidad de crímenes cometidos, diferenciados por sexo (femenino y masculino). Cada barra representa una categoría específica de crímenes con las cifras exactas de cada sexo. En la mayoría de las categorías, los crímenes cometidos a hombres son superiores a los cometidos a las mujeres, como podemos ver en aquellas donde la diferencia es particularmente significativa, como en la primera categoría (47,33 mil hombres vs. 33,27 mil mujeres) y la última categoría (16 mil hombres vs. 10,59 mil mujeres).



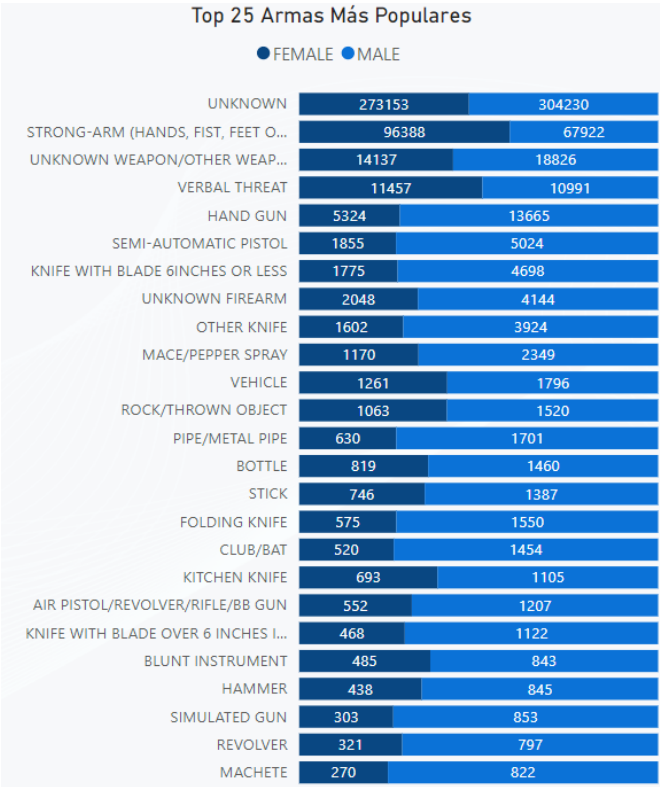
*Figura 19. Crímenes Por Sexo*

La *Figura 20* muestra las 10 edades más comunes de las personas involucradas en crímenes, expresadas en términos de cantidad y porcentaje del total. La edad de 30 años es la más frecuente entre las personas involucradas en crímenes, representando el 10.65% del total. La mayoría de los crímenes parecen estar concentrados en personas de edades entre 26 y 35 años, lo que puede indicar una tendencia específica en este grupo etario.



*Figura 20. Top 10 Edades Más Populares*

La *Figura 21*, revela que las armas más comunes entre hombres y mujeres son las manos, armas de fuego desconocidas, pistolas semiautomáticas y cuchillos. Se observan diferencias de género, con los hombres más propensos a ser víctimas de armas de fuego y las mujeres de armas no letales. Armas desconocidas y amenazas verbales también son relevantes.



*Figura 21. Top 25 Armas Más Populares*

#### 4. DISCUSIÓN (O ANÁLISIS DE RESULTADOS)

El análisis de los datos del comportamiento criminal en Los Ángeles entre 2020 y 2024 revela varios patrones significativos que permiten una comprensión más profunda de la dinámica delictiva en la ciudad. Los resultados principales se dividen en tres áreas clave: la temporalidad de los crímenes, la distribución geográfica de los incidentes y las características de los delitos y las víctimas.

El estudio identifica que los primeros trimestres de cada año presentan una mayor incidencia de crímenes, especialmente en 2021, 2022 y 2023. Este patrón puede estar relacionado con varios factores, incluidos los cambios estacionales, el comportamiento económico y social post-festividades, y la posible relajación de medidas de seguridad tras periodos de vigilancia intensificada. Este hallazgo sugiere que las estrategias de prevención del crimen deben intensificarse a principios de cada año para mitigar este aumento previsible en la actividad delictiva.

La identificación de zonas calientes de criminalidad mediante técnicas de visualización de datos muestra que ciertas áreas de la ciudad, históricamente conflictivas, siguen siendo focos de actividad criminal. Las áreas con alta densidad de población y aquellas con mayor actividad económica son particularmente vulnerables. La concentración de crímenes en estas zonas subraya la necesidad de políticas de seguridad específicas y la asignación de recursos policiales adicionales en estas áreas.

El análisis revela que las armas de fuego y cuchillos son las herramientas delictivas más comunes, lo que coincide con la preocupación generalizada sobre la violencia armada en las ciudades estadounidenses. Las armas de fuego, especialmente las pistolas semiautomáticas, son responsables de una proporción significativa de delitos violentos, mientras que los cuchillos están frecuentemente involucrados en crímenes como asaltos y robos. Además, las víctimas de crímenes presentan una variabilidad considerable en cuanto a edad, género y ascendencia, lo que indica que la violencia afecta a diversos segmentos de la población. No obstante, se observa una mayor incidencia de víctimas masculinas y de ascendencia hispana/latina, lo cual debe ser considerado en la formulación de políticas de intervención y apoyo a las víctimas.

Los resultados de este estudio son consistentes con investigaciones previas que han documentado patrones similares en otras grandes ciudades. Sin embargo, la aplicación de la metodología CRISP-DM en este análisis proporciona un enfoque más sistemático y basado en datos para la identificación de tendencias y la formulación de recomendaciones. La integración de técnicas avanzadas de minería de datos y análisis estadístico ha permitido una comprensión más granular y precisa de los patrones delictivos en Los Ángeles.

#### 5. CONCLUSIONES

El estudio concluye que hay patrones claros en el comportamiento criminal en Los Ángeles entre 2020 y 2024, con picos en los primeros trimestres de cada año y en ciertas zonas geográficas. Las armas de fuego y cuchillos son las más utilizadas en delitos. Estos hallazgos resaltan la importancia de desarrollar políticas públicas basadas en datos para mejorar la seguridad. Futuras investigaciones deberían centrarse en el análisis de factores socioeconómicos y su influencia en los patrones delictivos, así como en la evaluación de la efectividad de las intervenciones implementadas a partir de este estudio.

Se identificaron patrones claros de incremento de crímenes durante los primeros trimestres de cada año. Este hallazgo subraya la necesidad de reforzar las medidas de seguridad y vigilancia durante estos periodos específicos. Las autoridades deben planificar y asignar recursos adicionales para estos meses de alta incidencia delictiva, incluyendo campañas de concienciación comunitaria y programas de prevención del delito.

Las zonas calientes de criminalidad siguen siendo las mismas áreas históricamente conflictivas. Este resultado destaca la importancia de una vigilancia sostenida y la implementación de estrategias de intervención específicas en estas áreas. Las políticas de seguridad deben centrarse en la rehabilitación y revitalización de estas zonas, abordando no solo la criminalidad, sino también los factores subyacentes como la pobreza, la falta de oportunidades laborales y la deficiencia en servicios públicos.

Las armas de fuego, especialmente las pistolas semiautomáticas, y los cuchillos son las herramientas delictivas más comunes. Este hallazgo requiere una respuesta robusta en términos de control de armas y programas de desarme. Las políticas públicas deben incluir medidas estrictas de control de armas, programas educativos sobre los peligros de las armas y esfuerzos comunitarios para reducir la posesión ilegal de armas.

Las víctimas de crímenes varían significativamente en cuanto a edad, género y ascendencia, aunque se observa una mayor incidencia de víctimas masculinas y de ascendencia hispana/latina. Esto sugiere

---

la necesidad de programas de apoyo y protección específicos para estas comunidades, incluyendo servicios de asistencia a víctimas y programas de prevención dirigidos a jóvenes en riesgo.

Este estudio establece una base sólida para futuras investigaciones. Se sugiere expandir el análisis para incluir factores socioeconómicos y demográficos, así como evaluar la efectividad de las intervenciones implementadas. La integración de tecnologías avanzadas, como el análisis predictivo y la inteligencia artificial, podría mejorar aún más la capacidad de prevenir y responder a la criminalidad. La pandemia de COVID-19 ha tenido un impacto significativo en los patrones delictivos, alterando la dinámica social y económica de la ciudad. El estudio refleja cambios en la actividad delictiva durante los periodos de confinamiento y post-confinamiento, sugiriendo que las políticas de seguridad deben adaptarse a las nuevas realidades post-pandemia.

Concisamente, las conclusiones de este estudio destacan la importancia de una estrategia integral y basada en datos para abordar la criminalidad en Los Ángeles. La identificación de patrones temporales y geográficos, junto con un entendimiento detallado de las características de los delitos y las víctimas, proporciona una base sólida para la formulación de políticas efectivas y la asignación óptima de recursos. Además, subraya la necesidad de continuar investigando y adaptando las estrategias de seguridad a medida que evolucionan las dinámicas sociales y económicas de la ciudad.

## 6. REFERENCIAS BIBLIOGRÁFICAS

- [1] J. J. Espinosa-Zúñiga, "Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública," *Ing. Investig. y Tecnol.*, vol. 21, no. 1, pp. 1–13, Jan. 2020, doi: 10.22201/FI.25940732E.2020.21N1.008.
- [2] J. M. Moine, S. E. Gordillo, and A. S. Haedo, "Análisis comparativo de metodologías para la gestión de proyectos de minería de datos," 2011, Accessed: May 05, 2024. [Online]. Available: <http://sedici.unlp.edu.ar/handle/10915/18749>.