



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD CIENCIAS DE LA COMPUTACIÓN

**"ANÁLISIS DE PATRONES DE CRIMINALIDAD EN
LA CIUDAD DE LOS ANGELES (2020 A LA
ACTUALIDAD)"**

ABRAHAM FUENTES LÓPEZ

INTRODUCCIÓN A LA CIENCIA DE DATOS

M.C. JAIME ALEJANDRO ROMERO SIERRA

21/11/2024

Reporte Fase 1: Descubrimiento del Proyecto

1. Título del Proyecto

"Análisis de Patrones de Criminalidad en la Ciudad de Los Ángeles (2020 a la Actualidad)"

2. Objetivo del Proyecto

El presente proyecto tiene como objetivo analizar los patrones de criminalidad en la ciudad de Los Ángeles desde el año 2020 hasta la actualidad. La finalidad es proporcionar información que permita mejorar la asignación de recursos policiales, diseñar políticas de seguridad más eficaces y reducir la incidencia delictiva en áreas con mayor concentración de crímenes. Este análisis busca identificar tendencias, factores clave y relaciones entre las características de los delitos y las condiciones demográficas y geográficas de la ciudad.

3. Descripción del Problema

Desde 2020, la ciudad de Los Ángeles ha enfrentado un aumento significativo en la criminalidad, lo cual ha generado preocupación tanto en las autoridades como en las comunidades locales. A pesar del despliegue de esfuerzos para mitigar esta situación, la respuesta de las fuerzas del orden se ha visto obstaculizada por la falta de comprensión detallada sobre los patrones de criminalidad en términos de ubicación, tiempo y características demográficas. La actual distribución de recursos policiales se ha basado en enfoques reactivos que no consideran de manera precisa las áreas y momentos de mayor incidencia delictiva.

La ausencia de un análisis profundo de los factores que influyen en la frecuencia y tipo de delitos limita la capacidad de los responsables de seguridad para implementar estrategias preventivas adecuadas. Este proyecto busca identificar los patrones de criminalidad más relevantes en la ciudad, a fin de optimizar la asignación de recursos y mejorar las políticas de seguridad pública. Al abordar estos vacíos en el conocimiento, se espera contribuir a una reducción efectiva de la incidencia delictiva y al fortalecimiento de la seguridad en las comunidades más afectadas de Los Ángeles.

4. Recursos Disponibles

- **Tecnología y Herramientas:**
 - Python (Pandas, NumPy) para manipulación y análisis de datos.
 - Matplotlib y Seaborn para la visualización de datos.
- **Datos:**

- **Número_de_Reporte:** Identificador único del incidente.
- **Fecha_del_Reporte:** Fecha en que se presentó el reporte.
- **Fecha_del_Incidente:** Fecha del incidente.
- **Hora_del_Incidente:** Hora del incidente.
- **Área:** Ubicación general del incidente (aproximado por la cuadra).
- **Código_de_Delito:** Clasificación del tipo de crimen.
- **Edad_de_la_Víctima:** Edad de la persona afectada.
- **Sexo_de_la_Víctima:** Género de la víctima.
- **Descendencia_de_la_Víctima:** Origen étnico/racial de la víctima.
- **Lugar_del_Incidente:** Tipo de lugar donde ocurrió el delito (residencia, vía pública, etc.).
- **Arma_Utilizada:** Tipo de arma si es que se empleó alguna.
- **Estado_del_Caso:** Tipo de categoría del sospechoso (continúa la investigación, Desconocido, Adulto arrestado, etc).
- Latitud: Coordenadas de latitud de Los Angeles
- Longitud: Coordenadas de Longitud de los Angeles

5. Hipótesis Iniciales

- **Hipótesis 1:** Los delitos violentos tienen mayor prevalencia durante la noche y en áreas con baja presencia policial.
- **Hipótesis 2:** Las víctimas de delitos contra la propiedad tienden a ser personas mayores de 40 años en áreas residenciales.
- **Hipótesis 3:** Las áreas con mayor densidad poblacional presentan un número más alto de delitos reportados en comparación con áreas menos pobladas.

6. Definición de Stakeholders Clave

- **Departamento de Policía de Los Ángeles (LAPD):** Principal beneficiario del análisis; utilizarán los hallazgos para mejorar la asignación de recursos y ajustar estrategias de vigilancia.
- **Gobierno de Los Ángeles:** Podrán usar la información para el diseño e implementación de políticas de seguridad pública más eficientes.

- **Comunidades locales:** Indirectamente beneficiadas por las mejoras en la prevención del crimen y la mayor seguridad en sus áreas de residencia.
- **Investigadores y analistas de criminalidad:** Utilizarán los datos y análisis para estudios más profundos de comportamiento criminal.

7. Preguntas Clave

- ¿Cuáles son las áreas de Los Ángeles con mayor frecuencia de crímenes violentos?
- ¿Qué tipos de delitos son más comunes en cada área?
- ¿Existe una relación entre la hora del día y el tipo de crimen cometido?
- ¿Cuáles son las edades y géneros más afectados por delitos específicos?
- ¿Qué tipos de premisas son más propensas a ser escenario de ciertos delitos?
- ¿Cómo varían los patrones de criminalidad en función del origen étnico de las víctimas?
- ¿Hay picos en la actividad criminal durante ciertos meses del año?
- ¿Las áreas con mayor presencia policial tienen una menor incidencia delictiva?
- ¿Cómo afecta la densidad poblacional la frecuencia de delitos en diferentes áreas?
- ¿Qué impacto tienen las intervenciones recientes del LAPD en la reducción de criminalidad?

8. Fuentes de Datos Identificadas

- **Datos históricos de incidentes criminales del LAPD** (2020 - Actualidad).
- **Datos geográficos de áreas de criminalidad:** Definición de las zonas policiales en Los Ángeles.
- **Datos demográficos de Los Ángeles:** Información sobre la distribución de la población por edad, género y etnia.
- **Registros de intervenciones policiales recientes:** Información sobre el despliegue de recursos y programas comunitarios.

9. Justificación del Proyecto

La criminalidad es un tema de gran importancia para el bienestar social y económico de cualquier ciudad. Este análisis proporcionará información crucial que permitirá a las autoridades optimizar la distribución de recursos policiales y diseñar políticas de prevención de delitos más efectivas. Al comprender mejor las dinámicas de los incidentes delictivos —quiénes son las víctimas, dónde ocurren los delitos y en qué condiciones— se podrán tomar decisiones más fundamentadas para reducir el crimen, mejorar la seguridad pública y aumentar la calidad de vida de los ciudadanos.

10. Cantidad y Tipo de Datos

- **Cantidad de Datos:** La base de datos contiene aproximadamente 937,536 registros de incidentes reportados entre 2020 y la actualidad. Sin embargo, no se dispone de un registro completo del 100% de los reportes, lo que puede influir en el análisis.
- **Tipo de Datos:**
 - **Estructurados:** Incluyen identificadores de incidentes, fechas y horas, áreas de ocurrencia, códigos de delitos, y datos demográficos de las víctimas (edad, género, origen étnico). También se registran características del lugar y el uso de armas.
 - **Geoespaciales:** Datos que permiten realizar análisis geográficos y mapas de calor para identificar puntos críticos de criminalidad.
 - **Estados del sospechoso (Status):** Categorías como "Adult Arrest", "Invest Cont", "Adult Other", "Juv Arrest", "Juv Other" y "UNK", que reflejan el estado procesal de cada incidente.

Metodología

Análisis inicial:

Resumen estadístico de la base de datos antes de la limpieza

```
df = pd.read_csv('content/drive/MyDrive/df_socio (2).csv')
df
```

	BR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victim_age	Victim_sex	Victim_descent	Premis	Weapon	Status	LOCATION	LAT	LOW
0	190328475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wilshire	VEHICLE - STOLEN	0.0	M	O	STREET	NaN	Adult Arrest	NaN	34.0375	-118.3506
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	47.0	M	O	BUS STOP/LAYOVER (ALSO QUERY 124)	NaN	Invest Cont	1000 S FLOWER ST	34.0444	-118.2628
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	19.0	X	X	MULTIUNIT DWELLING (APARTMENT, DUPLEX, ETC)	NaN	Invest Cont	1400 W 37TH ST	invalid	-118.3002
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	19.0	NaN	O	CLOTHING STORE	NaN	Invest Cont	14000 RIVERSIDE DR	34.1576	-118.4387
4	NaN	08/18/2022 12:00:00 AM	08/17/2020 12:00:00 AM	1200.0	Hollywood	THEFT OF IDENTITY	28.0	M	H	SIDEWALK	NaN	Invest Cont	invalid	34.0944	-118.3277
...
1089038	200212691.0	07/17/2020 12:00:00 AM	NaN	130.0	Rampart	VANDALISM - FELONY (\$400 & OVER, ALL CHURCH VA...	0.0	X	X	HOSPITAL	NaN	Invest Cont	1200 WILSHIRE BL	34.053	-118.2649
1089039	220221781.0	12/16/2022 12:00:00 AM	12/16/2022 12:00:00 AM	1420.0	Rampart	ATTEMPTED ROBBERY	30.0	M	H	NaN	REVOLVER	Adult Arrest	8TH ST	34.0506	NaN
1089040	241110191.0	08/23/2024 12:00:00 AM	08/23/2024 12:00:00 AM	1530.0	Northeast	VEHICLE - STOLEN	0.0	NaN	NaN	STREET	NaN	Invest Cont	2600 N FIGUEROA ST	34.0884	-118.2258
1089041	220112932.0	05/23/2022 12:00:00 AM	05/23/2022 12:00:00 AM	1900.0	Central	BATTERY - SIMPLE ASSAULT	18.0	M	H	STREET	STRONG-ARM (HANDS, FIST, FEET OR BODY FORCE)	Invalid	5TH	34.0481	-118.2507
1089042	200320464.0	11/16/2020 12:00:00 AM	11/16/2020 12:00:00 AM	600.0	Southwest	VANDALISM - MISDEMEANOR (\$399 OR UNDER)	57.0	F	B	VEHICLE, PASSENGER/TRUCK	UNKNOWN WEAPON/OTHER WEAPON	Invest Cont	3400 EDGEHILL RD	34.0251	-118.3305

Nuestro Dataframe cuenta con un total de 1,089,043 filas y 15 columnas:

```
# Resumen estadístico de la base de datos antes de limpiarse

# Cantidad de filas y columnas respectivamente
df.shape
```

(1089043, 15)

Total de filas duplicadas:

```
# Total de filas duplicadas encontradas
df.duplicated().sum()
```

38269

Porcentaje de valores faltantes por columna:

```
# Porcentaje de valores faltantes por columna
df.isnull().mean()*100
```

	0
DR_NO	2.999973
Date_Reported	2.999973
Date_occured	2.999973
Time_occured	2.999973
Area	2.999973
Crime_Code	2.999973
Victim_age	2.999973
Victim_sex	16.520927
Victim_descent	16.535894
Premis	3.059383
Weapon	67.564274
Status	2.999973
LOCATION	2.999973
LAT	2.999973
LON	2.999973

Descripción de los tipos de datos originales:

```
#Descripción de los tipos de datos originales.

df.info()
```

#	Column	Non-Null Count	Dtype
0	DR_NO	1056372 non-null	float64
1	Date_Reported	1056372 non-null	object
2	Date_occured	1056372 non-null	object
3	Time_occured	1056372 non-null	float64
4	Area	1056372 non-null	object
5	Crime_Code	1056372 non-null	object
6	Victim_age	1056372 non-null	float64
7	Victim_sex	909123 non-null	object
8	Victim_descent	908960 non-null	object
9	Premis	1055725 non-null	object
10	Weapon	353239 non-null	object
11	Status	1056372 non-null	object
12	LOCATION	1056372 non-null	object
13	LAT	1056372 non-null	object
14	LON	1056372 non-null	object

dtypes: float64(3), object(12)
memory usage: 124.6+ MB

Proceso de limpieza de datos

Eliminación de las columnas de LOCATION, LAT Y LON

No las ocuparemos ya que se usan para espacios geográficos específicos y aún no aprendemos a usar esas herramientas.

Elimina las columnas de LOCATION, LAT Y LON
Los argumentos que se usan para espacios geográficos específicos y aún no aprendemos a usar esas herramientas
Ejecutar celda (Ctrl+Enter)
celda ejecutada desde el último cambio
ejecutada por Abraham Fuentes
2:08 (hace 1 hora)
se ha ejecutado en 0.474 s

	DR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victim_age	Victim_sex	Victim_descent	Premis	Weapon	Status
0	190326475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wilshire	VEHICLE - STOLEN	0.0	M	O	STREET	NaN	Adult Arrest
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	47.0	M	O	BUS STOP/LAYOVER (ALSO QUERY 124)	NaN	Invest Conf
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	19.0	X	X	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	NaN	Invest Conf
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	19.0	NaN	O	CLOTHING STORE	NaN	Invest Conf
4	NaN	08/18/2022 12:00:00 AM	08/17/2020 12:00:00 AM	1200.0	Hollywood	THEFT OF IDENTITY	28.0	M	H	SIDEWALK	NaN	Invest Conf
...
1089038	200212691.0	07/17/2020 12:00:00 AM	NaN	130.0	Rampart	VANDALISM - FELONY (\$400 & OVER, ALL CHURCH VA...	0.0	X	X	HOSPITAL	NaN	Invest Conf
1089039	220221781.0	12/16/2022 12:00:00 AM	12/16/2022 12:00:00 AM	1420.0	Rampart	ATTEMPTED ROBBERY	30.0	M	H	NaN	REVOLVER	Adult Arrest
1089040	241110191.0	08/23/2024 12:00:00 AM	08/23/2024 12:00:00 AM	1530.0	Northeast	VEHICLE - STOLEN	0.0	NaN	NaN	STREET	NaN	Invest Conf
1089041	220112932.0	05/23/2022 12:00:00 AM	05/23/2022 12:00:00 AM	1900.0	Central	BATTERY - SIMPLE ASSAULT	18.0	M	H	STREET	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	Invalid
1089042	200320464.0	11/16/2020 12:00:00 AM	11/16/2020 12:00:00 AM	600.0	Southwest	VANDALISM - MISDEAMEANOR (\$399 OR UNDER)	57.0	F	B	VEHICLE, PASSENGER/TRUCK	UNKNOWN WEAPON/OTHER WEAPON	Invest Conf

1089043 rows x 12 columns

Eliminación de las filas con NaN de ciertas columnas

df

	DR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victim
0	190326475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wilshire	VEHICLE - STOLEN	
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	
4	NaN	08/18/2022 12:00:00 AM	08/17/2020 12:00:00 AM	1200.0	Hollywood	THEFT OF IDENTITY	
...
1089038	200212691.0	07/17/2020 12:00:00 AM	NaN	130.0	Rampart	VANDALISM - FELONY (\$400 & OVER, ALL CHURCH VA...	
1089039	220221781.0	12/16/2022 12:00:00 AM	12/16/2022 12:00:00 AM	1420.0	Rampart	ATTEMPTED ROBBERY	
1089040	241110191.0	08/23/2024 12:00:00 AM	08/23/2024 12:00:00 AM	1530.0	Northeast	VEHICLE - STOLEN	
1089041	220112932.0	05/23/2022 12:00:00 AM	05/23/2022 12:00:00 AM	1900.0	Central	BATTERY - SIMPLE ASSAULT	
1089042	200320464.0	11/16/2020 12:00:00 AM	11/16/2020 12:00:00 AM	600.0	Southwest	VANDALISM - MISDEAMEANOR (\$399 OR UNDER)	

1089043 rows x 12 columns

#Elimina las filas con NaN de ciertas columnas
df = df.dropna(subset=['DR_NO'])
df = df.dropna(subset=['Date_Reported'])
df = df.dropna(subset=['Crime_Code'])
df

	DR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victi
0	190326475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wilshire	VEHICLE - STOLEN	
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	
5	231808869.0	04/04/23	12/01/20 0:00	2300.0	Southeast	THEFT OF IDENTITY	
...
1089038	200212691.0	07/17/2020 12:00:00 AM	NaN	130.0	Rampart	VANDALISM - FELONY (\$400 & OVER, ALL CHURCH VA...	
1089039	220221781.0	12/16/2022 12:00:00 AM	12/16/2022 12:00:00 AM	1420.0	Rampart	ATTEMPTED ROBBERY	
1089040	241110191.0	08/23/2024 12:00:00 AM	08/23/2024 12:00:00 AM	1530.0	Northeast	VEHICLE - STOLEN	
1089041	220112932.0	05/23/2022 12:00:00 AM	05/23/2022 12:00:00 AM	1900.0	Central	BATTERY - SIMPLE ASSAULT	
1089042	200320464.0	11/16/2020 12:00:00 AM	11/16/2020 12:00:00 AM	600.0	Southwest	VANDALISM - MISDEAMEANOR (\$399 OR UNDER)	

993923 rows x 12 columns

Cambio valores nulos a valores que podrían reemplazarlos

df.head()

	DR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victim_age	Victim_sex	Victim_descnt	Premis	Weapon	Status
0	190326475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wilshire	VEHICLE - STOLEN	0.0	M	O	STREET	NaN	Adult Arrest
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	47.0	M	O	BUS STOP/LAYOVER (ALSO QUERY 124)	NaN	Invest Cont
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	19.0	X	X	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	NaN	Invest Cont
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	19.0	NaN	O	CLOTHING STORE	NaN	Invest Cont
5	231808869.0	04/04/23	12/01/20 0:00	2300.0	Southeast	THEFT OF IDENTITY	41.0	M	H	SINGLE FAMILY DWELLING	NaN	NaN

#Cambio valores nulos a valores que podrían reemplazarlos
df['Date_occured'] = df['Date_occured'].fillna('DESCONOCIDO')
df['Victim_descnt'] = df['Victim_descnt'].fillna('DESCONOCIDO')
df['Weapon'] = df['Weapon'].fillna('SIN_AGRESION')
df.head()

	DR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victim_age	Victim_sex	Victim_descnt	Premis	Weapon	Status
0	190326475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wilshire	VEHICLE - STOLEN	0.0	M	O	STREET	SIN_AGRESION	Adult Arrest
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	47.0	M	O	BUS STOP/LAYOVER (ALSO QUERY 124)	SIN_AGRESION	Invest Cont
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	19.0	X	X	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	SIN_AGRESION	Invest Cont
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	19.0	NaN	O	CLOTHING STORE	SIN_AGRESION	Invest Cont
5	231808869.0	04/04/23	12/01/20 0:00	2300.0	Southeast	THEFT OF IDENTITY	41.0	M	H	SINGLE FAMILY DWELLING	SIN_AGRESION	NaN

Imputación de valores faltantes

df

	DR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victim_age	Victim_sex	Victim_descent	Premis	Weapon	Status
0	190326475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wilshire	VEHICLE - STOLEN	0.0	M	O	STREET	SIN_AGRESION	Adult Arrest
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	47.0	M	O	BUS STOP/LAYOVER (ALSO QUERY 124)	SIN_AGRESION	Invest Cort
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	19.0	X	X	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	SIN_AGRESION	Invest Cort
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	19.0	NaN	O	CLOTHING STORE	SIN_AGRESION	Invest Cort
5	231808869.0	04/04/23	12/01/20 0:00	2300.0	Southeast	THEFT OF IDENTITY	41.0	M	H	SINGLE FAMILY DWELLING	SIN_AGRESION	NaN
...
1089038	200212691.0	07/17/2020 12:00:00 AM	DESCONOCIDO	130.0	Rampart	VANDALISM - FELONY (\$400 & OVER, ALL CHURCH VA...	0.0	X	X	HOSPITAL	SIN_AGRESION	Invest Cort
1089039	220221781.0	12/16/2022 12:00:00 AM	12/16/2022 12:00:00 AM	1420.0	Rampart	ATTEMPTED ROBBERY	30.0	M	H	NaN	REVOLVER	Adult Arrest
1089040	241110191.0	08/23/2024 12:00:00 AM	08/23/2024 12:00:00 AM	1530.0	Northeast	VEHICLE - STOLEN	0.0	NaN	DESCONOCIDO	STREET	SIN_AGRESION	Invest Cort
1089041	220112932.0	05/23/2022 12:00:00 AM	05/23/2022 12:00:00 AM	1900.0	Central	BATTERY - SIMPLE ASSAULT	18.0	M	H	STREET	STRONG ARM (HANDS, FIST, FEET OR BODY FORCE)	Invalid
1089042	200320464.0	11/16/2020 12:00:00 AM	11/16/2020 12:00:00 AM	600.0	Southwest	VANDALISM - MISDEMEANOR (\$399 OR UNDER)	57.0	F	B	VEHICLE, PASSENGER/TRUCK	UNKNOWN WEAPON/OTHER WEAPON	Invest Cort

993923 rows x 12 columns

Imputacion de valores faltantes

```
# Victim_age -> media
# Time_occured -> media

# Area -> Moda
# Victim_sex -> Moda
# Premis -> Moda
# Status -> Moda

df['Victim_age'] = df['Victim_age'].fillna(df['Victim_age'].mean())
df['Time_occured'] = df['Time_occured'].fillna(df['Time_occured'].mean())

df['Area'] = df['Area'].fillna(df['Area'].mode()[0])
df['Victim_sex'] = df['Victim_sex'].fillna(df['Victim_sex'].mode()[0])
df['Premis'] = df['Premis'].fillna(df['Premis'].mode()[0])
df['Status'] = df['Status'].fillna(df['Status'].mode()[0])
df
```

df.isnull().sum()

	0
DR_NO	0
Date_Reported	0
Date_occured	0
Time_occured	0
Area	0
Crime_Code	0
Victim_age	0
Victim_sex	0
Victim_descent	0
Premis	0
Weapon	0
Status	0

Eliminación de las filas duplicadas

```
# total de filas duplicadas
df.duplicated().sum()

56387

# Eliminamos las filas duplicadas
df=df.drop_duplicates()
df
```

	DR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victim_age	Victim_sex	Victim_descent	Premis	Weapon	Status
0	190326475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wilshire	VEHICLE - STOLEN	0.0	M	O	STREET	SIN_AGRESION	Adult Arrest
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	47.0	M	O	BUS STOP/LAYOVER (ALSO QUERY 124)	SIN_AGRESION	Invest Cort
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	19.0	X	X	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	SIN_AGRESION	Invest Cort
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING GRAND THEFT (\$950.01 & OVER)	19.0	M	O	CLOTHING STORE	SIN_AGRESION	Invest Cort
5	231808869.0	04/04/23	12/01/20 0:00	2300.0	Southeast	THEFT OF IDENTITY	41.0	M	H	SINGLE FAMILY DWELLING	SIN_AGRESION	Invest Cort
...
1089034	231115828.0	11/15/2023 12:00:00 AM	11/12/23 0:00	2200.0	Northeast	VEHICLE - STOLEN	0.0	M	DESCONOCIDO	STREET	SIN_AGRESION	Invest Cort
1089038	200212691.0	07/17/2020 12:00:00 AM	DESCONOCIDO	130.0	Rampart	VANDALISM - FELONY (\$400 & OVER, ALL CHURCH VA..	0.0	X	X	HOSPITAL	SIN_AGRESION	Invest Cort
1089039	220221781.0	12/16/2022 12:00:00 AM	12/16/2022 12:00:00 AM	1420.0	Rampart	ATTEMPTED ROBBERY	30.0	M	H	STREET	REVOLVER	Adult Arrest
1089041	220112932.0	05/23/2022 12:00:00 AM	05/23/2022 12:00:00 AM	1900.0	Central	BATTERY - SIMPLE ASSAULT	18.0	M	H	STREET	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	Invalid
1089042	200320464.0	11/16/2020 12:00:00 AM	11/16/2020 12:00:00 AM	600.0	Southwest	VANDALISM - MISDEMEANOR (\$399 OR UNDER)	57.0	F	B	VEHICLE, PASSENGER/TRUCK	UNKNOWN WEAPON/OTHER WEAPON	Invest Cort

```
# total de filas duplicadas
df.duplicated().sum()

0
```

Verificación y eliminación de invalid values

```
# Verificando si tenemos invalid values
for i in lista_col:
    print(f"En la columna {i} los invalid_value son: {df[df[i] == 'bbb'].shape[0]}")

# NO tenemos invalid_value
```

```
En la columna DR_NO los invalid_value son: 0
En la columna Date_Reported los invalid_value son: 0
En la columna Date_occured los invalid_value son: 0
En la columna Time_occured los invalid_value son: 0
En la columna Area los invalid_value son: 0
En la columna Crime_Code los invalid_value son: 0
En la columna Victim_age los invalid_value son: 0
En la columna Victim_sex los invalid_value son: 0
En la columna Victim_descent los invalid_value son: 0
En la columna Premis los invalid_value son: 0
En la columna Weapon los invalid_value son: 0
En la columna Status los invalid_value son: 0
```

Resultados de datos

DataFrame final

	DR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victim_age	Victim_sex	Victim_descent	Premis	Weapon	Status
0	190326475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wilshire	VEHICLE - STOLEN	0.0	M	O	STREET	SIN_AGRESION	Adult Arrest
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	47.0	M	O	BUS STOP/LAYOVER (ALSO QUERY 124)	SIN_AGRESION	Invest Cont
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	19.0	X	X	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	SIN_AGRESION	Invest Cont
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	19.0	M	O	CLOTHING STORE	SIN_AGRESION	Invest Cont
5	231808869.0	04/04/23	12/01/20 0:00	2300.0	Southeast	THEFT OF IDENTITY	41.0	M	H	SINGLE FAMILY DWELLING	SIN_AGRESION	Invest Cont
...
1089934	231115828.0	11/15/2023 12:00:00 AM	11/12/23 0:00	2200.0	Northeast	VEHICLE - STOLEN	0.0	M	DESCONOCIDO	STREET	SIN_AGRESION	Invest Cont
1089938	200212691.0	07/17/2020 12:00:00 AM	DESCONOCIDO	130.0	Rampart	VANDALISM - FELONY (\$400 & OVER, ALL CHURCH VA...	0.0	X	X	HOSPITAL	SIN_AGRESION	Invest Cont
1089939	220221781.0	12/16/2022 12:00:00 AM	12/16/2022 12:00:00 AM	1420.0	Rampart	ATTEMPTED ROBBERY	30.0	M	H	STREET	REVOLVER	Adult Arrest
1089941	220112932.0	05/23/2022 12:00:00 AM	05/23/2022 12:00:00 AM	1900.0	Central	BATTERY - SIMPLE ASSAULT	18.0	M	H	STREET	STRONG-ARM (HANDS, FIST, FEET OR BODYLY FORCE)	Invalid
1089942	200320464.0	11/16/2020 12:00:00 AM	11/16/2020 12:00:00 AM	600.0	Southwest	VANDALISM - MISDEMEANOR (\$399 OR UNDER)	57.0	F	B	VEHICLE, PASSENGER/TRUCK	UNKNOWN WEAPON/OTHER WEAPON	Invest Cont

Nos quedan 937,536 Filas y 12 Columnas

```
df.info()

<class 'pandas.core.frame.DataFrame'>
Index: 942531 entries, 0 to 1089042
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   DR_NO                 942531 non-null float64
 1   Date_Reported         942531 non-null object  
 2   Date_occured          942531 non-null object  
 3   Time_occured          942531 non-null float64
 4   Area                  942531 non-null object  
 5   Crime_Code            942531 non-null object  
 6   Victim_age            942531 non-null float64
 7   Victim_sex            942531 non-null object  
 8   Victim_descent        942531 non-null object  
 9   Premis                942531 non-null object  
10   Weapon                942531 non-null object  
11   Status                942531 non-null object  
dtypes: float64(3), object(9)
memory usage: 93.5+ MB
```

Quedan 3 columnas de tipo float (numérico)

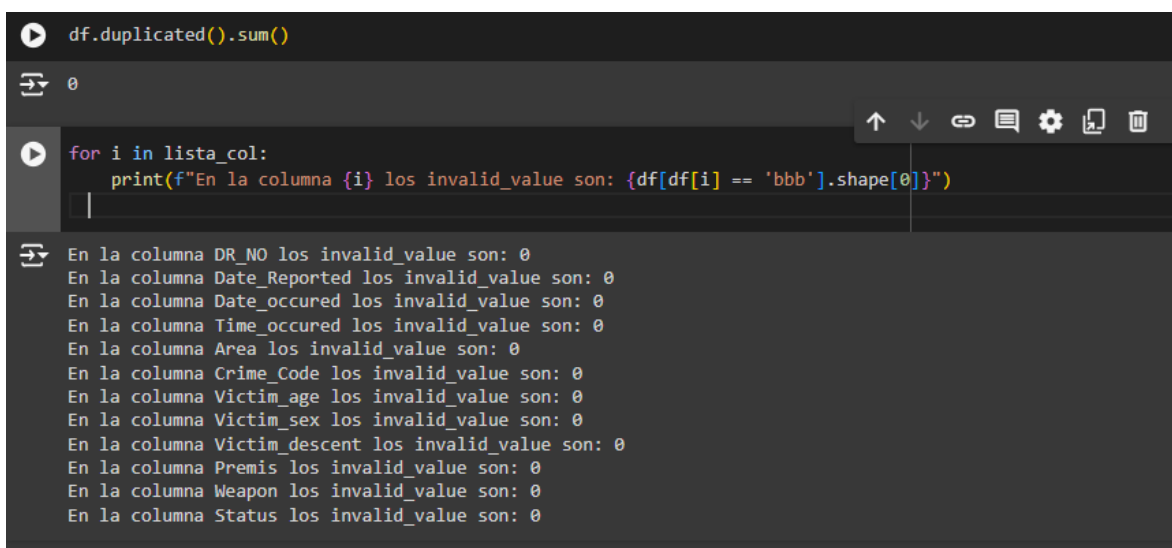
Tabla que muestre el porcentaje de valores faltantes final por columna.



```
df.isnull().mean()*100
```

	0
DR_NO	0.0
Date_Reported	0.0
Date_occured	0.0
Time_occured	0.0
Area	0.0
Crime_Code	0.0
Victim_age	0.0
Victim_sex	0.0
Victim_descent	0.0
Premis	0.0
Weapon	0.0
Status	0.0

Comprobación de que no hay duplicados ni valores inválidos.



```
df.duplicated().sum()
```

```
0
```

```
for i in lista_col:
    print(f"En la columna {i} los invalid_value son: {df[df[i] == 'bbb'].shape[0]}")
```

```
En la columna DR_NO los invalid_value son: 0
En la columna Date_Reported los invalid_value son: 0
En la columna Date_occured los invalid_value son: 0
En la columna Time_occured los invalid_value son: 0
En la columna Area los invalid_value son: 0
En la columna Crime_Code los invalid_value son: 0
En la columna Victim_age los invalid_value son: 0
En la columna Victim_sex los invalid_value son: 0
En la columna Victim_descent los invalid_value son: 0
En la columna Premis los invalid_value son: 0
En la columna Weapon los invalid_value son: 0
En la columna Status los invalid_value son: 0
```

Reporte Final

1. Análisis Exploratorio de Datos (EDA)

Importación de bibliotecas que se utilizarán, cada una con diferente función.

```
# Importación de bibliotecas

import pandas as pd # Pandas: Se utiliza para la manipulación y análisis de datos. Trabaja principalmente con estructuras como DataFrames.
import matplotlib.pyplot as plt # Matplotlib: Permite crear gráficos estáticos como líneas, barras, dispersión, entre otros.
import numpy as np # NumPy: Biblioteca para realizar operaciones numéricas, especialmente con arrays y matrices multidimensionales.
import seaborn as sns # Seaborn: Basado en Matplotlib, se usa para crear gráficos estadísticos atractivos y fáciles de interpretar.
import plotly.express as px # Plotly Express: Biblioteca para generar gráficos interactivos de manera rápida y sencilla.
```

Cargamos nuestra base de datos original y también cargamos nuestra base de datos que limpiamos anteriormente, esto para recuperar las columnas de Latitud y Longitud que borramos anteriormente, con esto nuestro df con el que trabajaremos ahora tiene la Latitud y Longitud, las cuales usaremos más adelante.

```
df2 = pd.read_csv('/content/Crime_Data_from_2020_to_Present.csv')

df = pd.read_csv('/content/Base_Limpia (2).csv')

df.head()
```

	DR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victim_age	Victim_sex	Victim_descent	Premis	Weapon	Status
0	190326475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wiltshire	VEHICLE - STOLEN	0.0	M	O	STREET	SIN_AGRESION	Adult Arrest
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	47.0	M	O	BUS STOP/LAYOVER (ALSO QUERY 124)	SIN_AGRESION	Invest Cont
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	19.0	X	X	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	SIN_AGRESION	Invest Cont
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	19.0	M	O	CLOTHING STORE	SIN_AGRESION	Invest Cont
4	231808869.0	04/04/23	12/01/20 0:00	2300.0	Southeast	THEFT OF IDENTITY	41.0	M	H	SINGLE FAMILY DWELLING	SIN_AGRESION	Invest Cont

```
# Recuperar las columnas LAT y LON desde df1 en base a DR_NO
df = df.merge(df2[['DR_NO', 'LAT', 'LON']], on='DR_NO', how='left')

df.head()
```

	DR_NO	Date_Reported	Date_occured	Time_occured	Area	Crime_Code	Victim_age	Victim_sex	Victim_descent	Premis	Weapon	Status	LAT	LON
0	190326475.0	03/01/20 0:00	03/01/20 0:00	2130.0	Wiltshire	VEHICLE - STOLEN	0.0	M	O	STREET	SIN_AGRESION	Adult Arrest	34.0375	-118.3506
1	200106753.0	02/09/20 0:00	02/08/20 0:00	1800.0	Central	BURGLARY FROM VEHICLE	47.0	M	O	BUS STOP/LAYOVER (ALSO QUERY 124)	SIN_AGRESION	Invest Cont	34.0444	-118.2628
2	200320258.0	11/11/20 0:00	11/04/20 0:00	1700.0	Southwest	BIKE - STOLEN	19.0	X	X	MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)	SIN_AGRESION	Invest Cont	34.0210	-118.3002
3	200907217.0	05/10/23 0:00	03/10/20 0:00	2037.0	Van Nuys	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	19.0	M	O	CLOTHING STORE	SIN_AGRESION	Invest Cont	34.1576	-118.4387

Renombramos las columnas de nuestro df para una mejor visualización

```
# Renombramos las columnas del df
df.rename(columns={
    'DR_NO': 'Número_de_Reporte',
    'Date_Reported': 'Fecha_del_Reporte',
    'Date_occured': 'Fecha_del_Incidente',
    'Time_occured': 'Hora_del_Incidente',
    'Area': 'Área',
    'Crime_Code': 'Código_de_Delito',
    'Victim_age': 'Edad_de_la_Víctima',
    'Victim_sex': 'Sexo_de_la_Víctima',
    'Victim_descent': 'Descendencia_de_la_Víctima',
    'Premis': 'Lugar_del_Incidente',
    'Weapon': 'Arma_Utilizada',
    'Status': 'Estado_del_Caso',
    'LAT': 'Latitud',
    'LON': 'Longitud'
}, inplace=True)
```

- Redondeamos edades
- Filtramos edades para que solo queden entre 0 y 100
- E imputamos edades de 0 a valores aleatorios (esto porque aparecen demasiados valores en 0)

```
# Redondear edades
df['Edad_de_la_Víctima'] = df['Edad_de_la_Víctima'].round()

# Filtrar solo edades entre 0 y 100
df = df[(df['Edad_de_la_Víctima'] >= 0) & (df['Edad_de_la_Víctima'] <= 100)]

# Imputa valores que son 0 a valores aleatorios, ya que aparecían demasiados en 0
indices_cero = df[df['Edad_de_la_Víctima'] == 0].index
edades_validas = df[df['Edad_de_la_Víctima'] > 0]['Edad_de_la_Víctima']
valores_imputados = np.random.choice(edades_validas, size=len(indices_cero))
df.loc[indices_cero, 'Edad_de_la_Víctima'] = valores_imputados
```

Cambiamos los nombres dentro de las columnas Descendencia_de_la_Víctima y Estado_del_Caso, esto para una mejor comprensión.

```
# Diccionario de mapeo para estandarizar valores
mapping = {
    'O': 'Otro',
    'X': 'Desconocido',
    'H': 'Hispano/Latino',
    'B': 'Negro/Afroamericano',
    'W': 'Blanco',
    'A': 'Asiático',
    'invalid': 'Desconocido',
    'DESCONOCIDO': 'Desconocido',
    '-': 'Desconocido',
    'K': 'Coreano',
    'C': 'Chino',
    'J': 'Japonés',
    'F': 'Filipino',
    'I': 'Indio Americano',
    'V': 'Vietnamita',
    'S': 'Samoano',
    'P': 'Polinesio',
    'Z': 'Otros Asiáticos',
    'G': 'Guamaniano',
    'U': 'Hawaiano',
    'D': 'Camboyano',
    'L': 'Laosiano'
}

# Reemplazar valores
df['Descendencia_de_la_Víctima'] = df['Descendencia_de_la_Víctima'].map(mapping)

# Diccionario para estandarizar valores
Estado_del_Caso_mapping = {
    'Adult Arrest': 'Arresto Adulto',
    'Invest Cont': 'Investigación en Curso',
    'Adult Other': 'Otro Adulto',
    'Juv Arrest': 'Arresto Juvenil',
    'Juv Other': 'Otro Juvenil',
    'UNK': 'Desconocido',
    'invalid': 'Desconocido'
}
```

1. Descripción general de los datos

- Filas: 937,310
- Columnas: 14

```
#Resumen del DataFrame
print(f"Filas: {df.shape[0]}, Columnas: {df.shape[1]}")
```

```
Filas: 937310, Columnas: 14
```

```
# Información general del DataFrame
print("Información general:")
df.info()
```

```
Información general:
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 937310 entries, 0 to 937535
```

```
Data columns (total 14 columns):
```

#	Column	Non-Null Count	Dtype
0	Número_de_Reporte	937310 non-null	float64
1	Fecha_del_Reporte	937310 non-null	object
2	Fecha_del_Incidente	937310 non-null	object
3	Hora_del_Incidente	937310 non-null	float64
4	Área	937310 non-null	object
5	Código_de_Delito	937310 non-null	object
6	Edad_de_la_Victima	937310 non-null	float64
7	Sexo_de_la_Victima	937310 non-null	object
8	Descendencia_de_la_Victima	937310 non-null	object
9	Lugar_del_Incidente	937310 non-null	object
10	Arma_Utilizada	937310 non-null	object
11	Estado_del_Caso	937310 non-null	object
12	Latitud	937310 non-null	float64
13	Longitud	937310 non-null	float64

```
dtypes: float64(5), object(9)
```

```
memory usage: 107.3+ MB
```


- Resumen estadístico por columnas numéricas
- Número de valores únicos por columna

```
# Resumen estadístico para columnas numéricas
print("\nResumen estadístico:")
print(df.describe())
```

Resumen estadístico:

	Número_de_Reporte	Hora_del_Incidente	Edad_de_la_Víctima \
count	9.373100e+05	937310.000000	937310.000000
mean	2.195663e+08	1338.936908	39.092323
std	1.285229e+07	640.911413	15.409292
min	2.113000e+03	1.000000	2.000000
25%	2.106061e+08	919.000000	28.000000
50%	2.208089e+08	1400.000000	36.000000
75%	2.309064e+08	1845.000000	49.000000
max	2.499253e+08	2359.000000	99.000000

	Latitud	Longitud
count	937310.000000	937310.000000
mean	33.994256	-118.077245
std	1.651480	5.724289
min	0.000000	-118.667600
25%	34.014600	-118.430300
50%	34.058900	-118.322500
75%	34.164900	-118.273900
max	34.334300	0.000000

```
# Número de valores únicos por columna
print("\nNúmero de valores únicos por columna:")
print(df.nunique())
```

Número de valores únicos por columna:

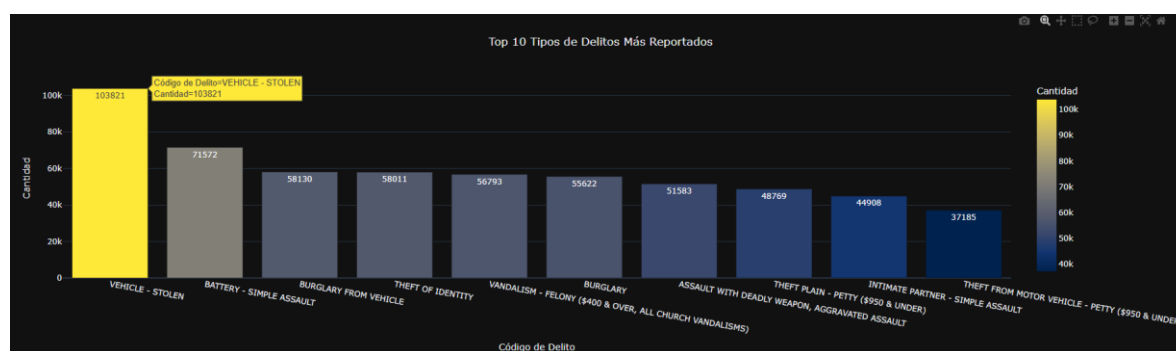
Número_de_Reporte	897882
Fecha_del_Reporte	1708
Fecha_del_Incidente	1708
Hora_del_Incidente	1440
Área	22
Código_de_Delito	139
Edad_de_la_Víctima	98
Sexo_de_la_Víctima	3
Descendencia_de_la_Víctima	19
Lugar_del_Incidente	306
Arma_Utilizada	80
Estado_del_Caso	6
Latitud	5409
Longitud	4973

2. Análisis de los Delitos

2.1 Distribución de los 10 Principales Códigos de Delito

- **Título:** Top 10 Tipos de Delitos Más Reportados
- **Descripción:** Muestra cuáles son los delitos más frecuentes y compara sus proporciones.

Este gráfico destaca los 10 delitos más reportados en Los Ángeles. El delito más común es **Vehículo robado** que representa 103821 casos, seguido de otros delitos como **Agresión física - agresión simple** y **Robo en vehículo**. Esto sugiere un enfoque en estos tipos de incidentes para estrategias preventivas.

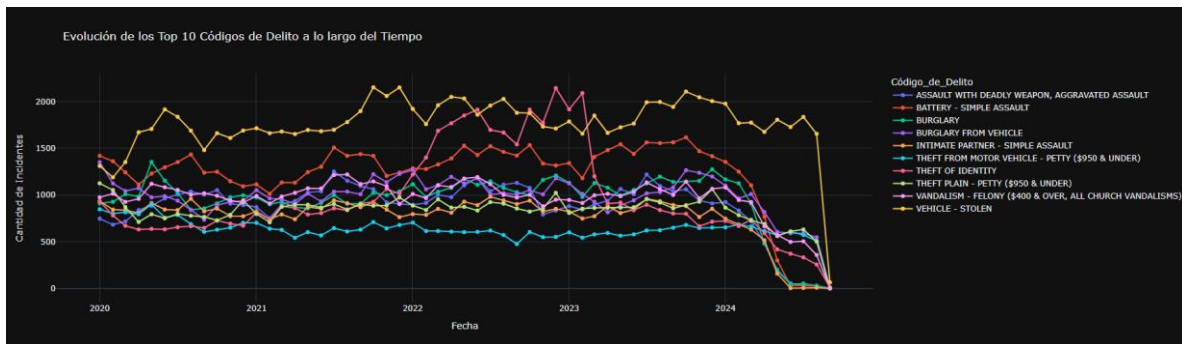


- Esto nos dice que la mayoría de los delitos se efectúan en la calle y en especial, los delitos se enfocan en vehículos, ya sea, el robo total de vehículo, por partes o de otro tipo.

2.2 Evolución de los Top 10 Códigos de Delito a lo Largo del Tiempo

- **Título:** Tendencia de los Principales Delitos con el Tiempo
- **Descripción:** Analiza cómo han cambiado las frecuencias de los delitos en el tiempo.

Este gráfico revela patrones temporales en los delitos más frecuentes durante el 2020 a la actualidad. Notamos gravemente la presencia de robo de vehículos durante los 4 años y también vemos como entre 2022 y 2023 hubo picos de Robos de identidad. Algunos otros se han mantenido relativamente constantes, pero todos empezaron a disminuir entre febrero y marzo del 2024

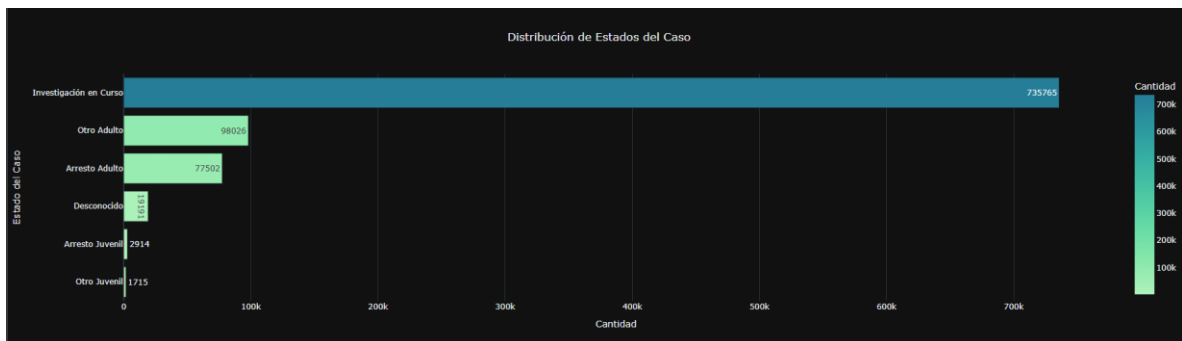


3. Estado de los Casos

3.1 Distribución de Estados del Caso

- **Título:** Estatus Actual de los Casos
- **Descripción:** Indica el progreso de los casos en el sistema legal.

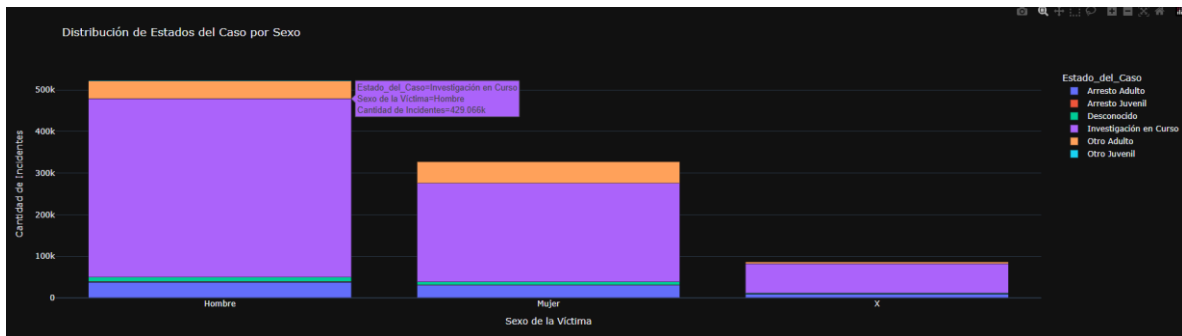
La mayoría de los casos reportados están clasificados como "Investigación en curso" siendo 735,765 lo que es un altísimo pico. Solo un pequeño porcentaje ha llegado a resolución definitiva. Esto sugiere posibles problemas de saturación o retrasos en los procesos.



3.2 Distribución de Estados del Caso por Sexo

- **Título:** Estatus del Caso por Género de la Víctima
- **Descripción:** Analiza si hay diferencias en cómo avanzan los casos según el género de la víctima.

Este gráfico muestra como los 3 géneros que aparecen, presentan una proporción similar en cuanto al estatus de los casos, esto quiere decir que no hay gran diferencia entre hombres, mujeres y X para la resolución de casos

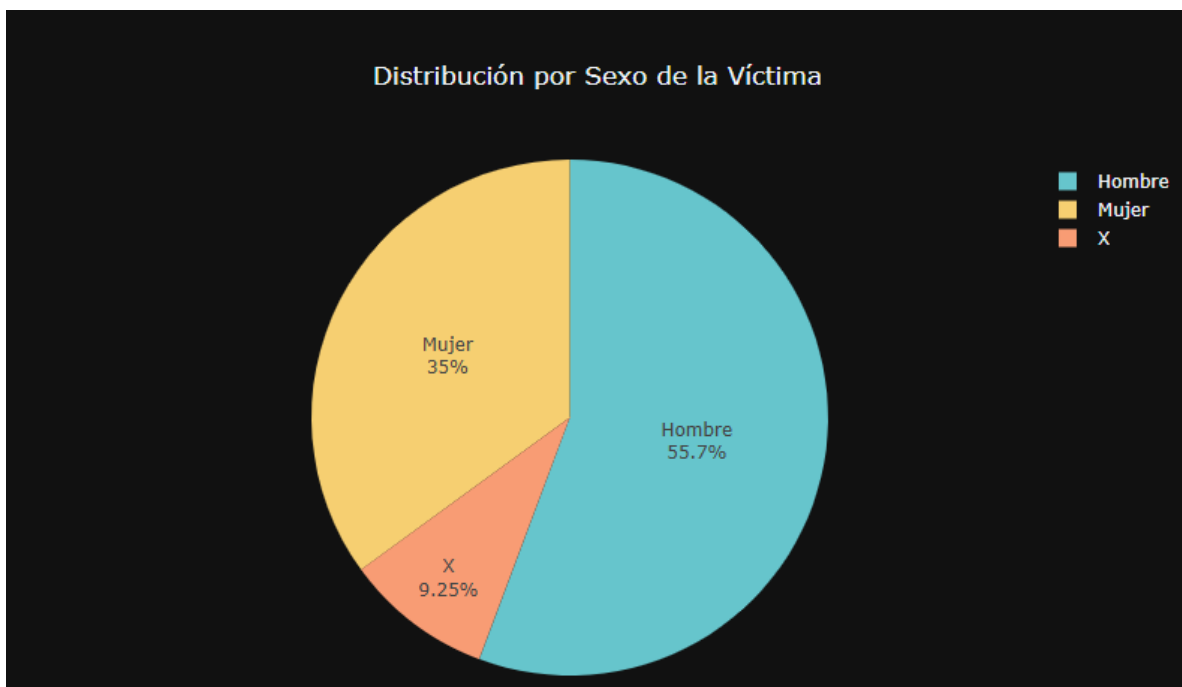


4. Caracterización de las Víctimas

4.1 Distribución por Sexo de la Víctima

- **Título:** Distribución por Género de las Víctimas
- **Descripción:** Muestra la proporción de hombres, mujeres, y otras identidades entre las víctimas.

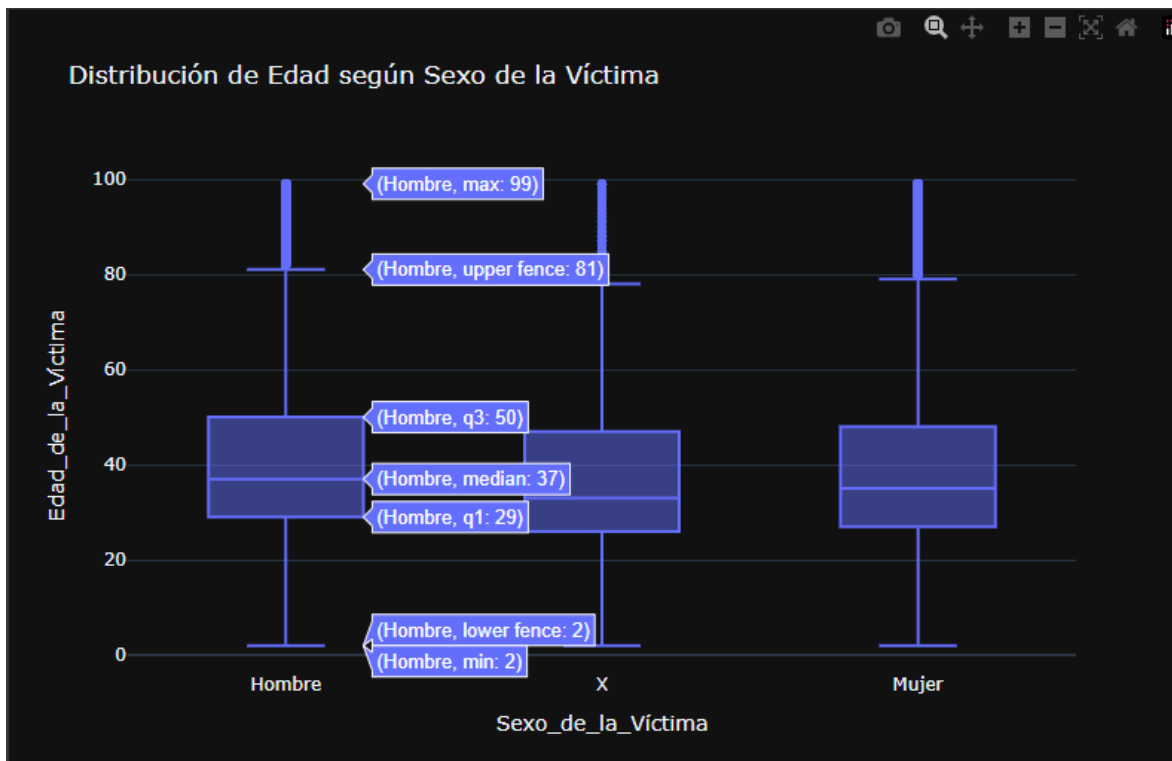
La mayoría de las víctimas son hombres con un 55.7%, seguidos por mujeres con un 35% y personas de género no especificado con 9.25%. Esto podría reflejar diferencias en la exposición o en la probabilidad de reportar incidentes. Y también muestra que pueden llegar a ser mas vulnerables los hombres a robos, accidentes o cualquier otro tipo de incidente



4.2 Distribución de Edad según Sexo de la Víctima

- **Título:** Distribución de la Edad por Género
- **Descripción:** Explora si la edad de las víctimas varía según el género.

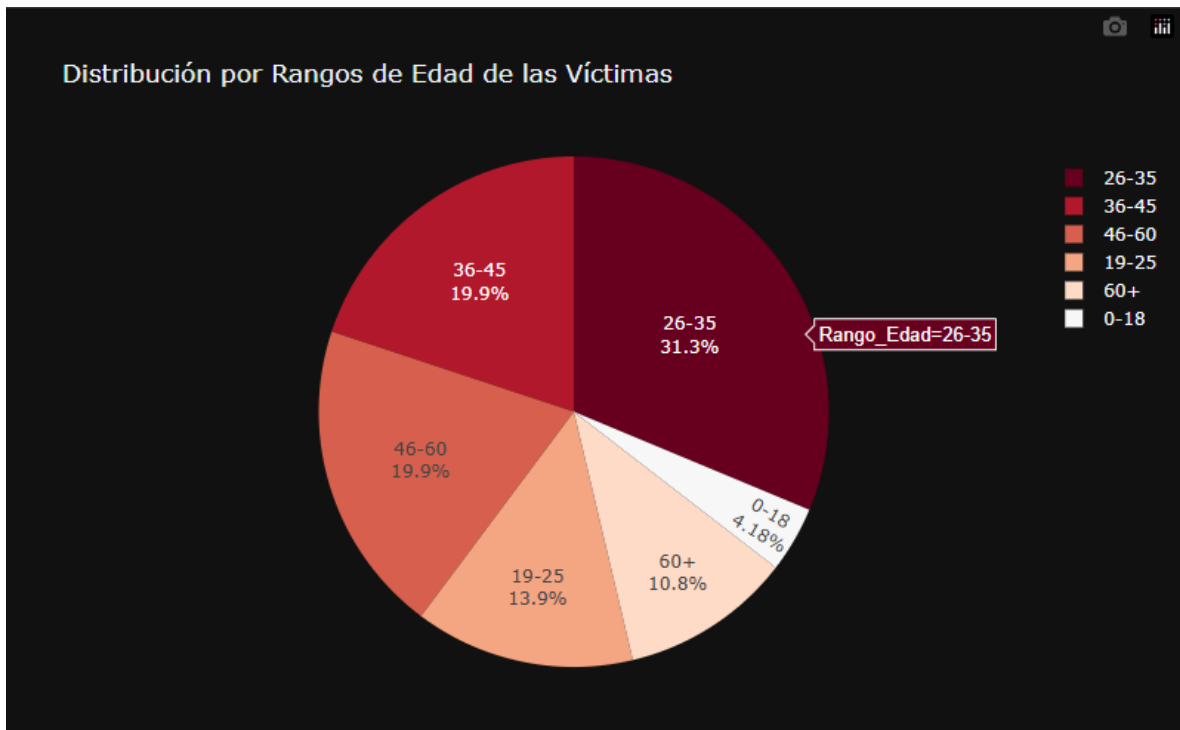
Este análisis revela en aproximación los 3 géneros se encuentran en un rango de entre 25 y 50, quiere decir que son mas propensos a tener un incidente



4.3 Distribución por Rangos de Edad de las Víctimas

- **Título:** Grupos Etarios Más Afectados
- **Descripción:** Identifica qué rangos de edad tienen mayor incidencia.

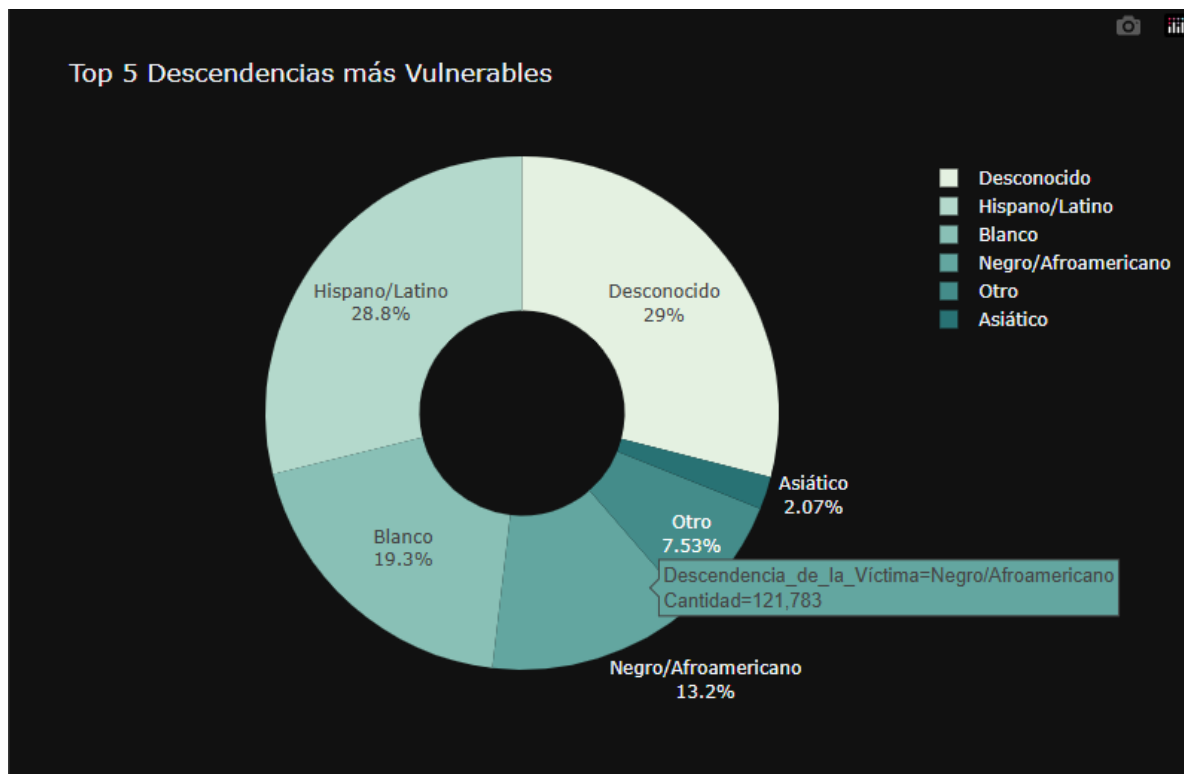
El rango de edad de 26 a 35 años es el más afectado, representando el 31.3% del total de víctimas. Esto podría estar relacionado con una combinación de factores socioeconómicos, mayor exposición a riesgos en la vida adulta temprana, y la violencia urbana asociada con la pobreza, el crimen organizado, y las pandillas.



4.4 Top 5 Descendencias Más Vulnerables

- **Título:** Descendencias con Mayor Frecuencia de Incidentes
- **Descripción:** Analiza las descendencias más comunes entre las víctimas.

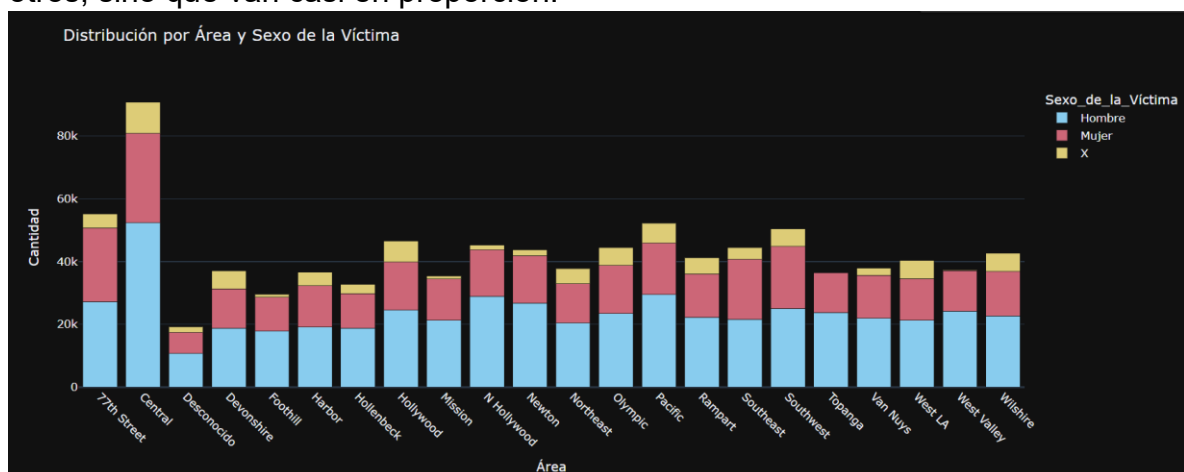
Las descendencias más reportadas incluyen **Desconocido** (esto tal vez a que no se hizo un registro apropiado), **Hispano/Latino**, y **Blancos**. Esto puede reflejar diferencias en la población o en factores sociales que afectan su vulnerabilidad. Los **Hispanos/Latinos** en Los Ángeles tienden a estar más representados en áreas con mayores desafíos socioeconómicos, como pobreza y desempleo, lo que aumenta su vulnerabilidad a delitos. Además, enfrentan barreras lingüísticas y miedo a la deportación, lo que dificulta su acceso a la justicia y a la denuncia de crímenes. Por otro lado, los **Blancos** suelen estar en áreas con más recursos, lo que podría explicar una menor vulnerabilidad a los delitos en comparación con otros grupos marginados.



4.5 Distribución por Área y Sexo de la Víctima

- **Título:** Análisis de Género por Área
- **Descripción:** Examina la distribución de las víctimas según su género y las áreas donde ocurrieron los incidentes.

Este gráfico muestra cómo se distribuyen los géneros de las víctimas en distintas áreas geográficas. Además, muestra como en **Central** y en **77th Street** hay mayor concentración de delitos, por lo que son zonas más riesgosas. No hay una notoria diferencia entre áreas que muestre que ciertos géneros sean mas vulnerables que otros, sino que van casi en proporción.



5. Análisis Espacio-Temporal

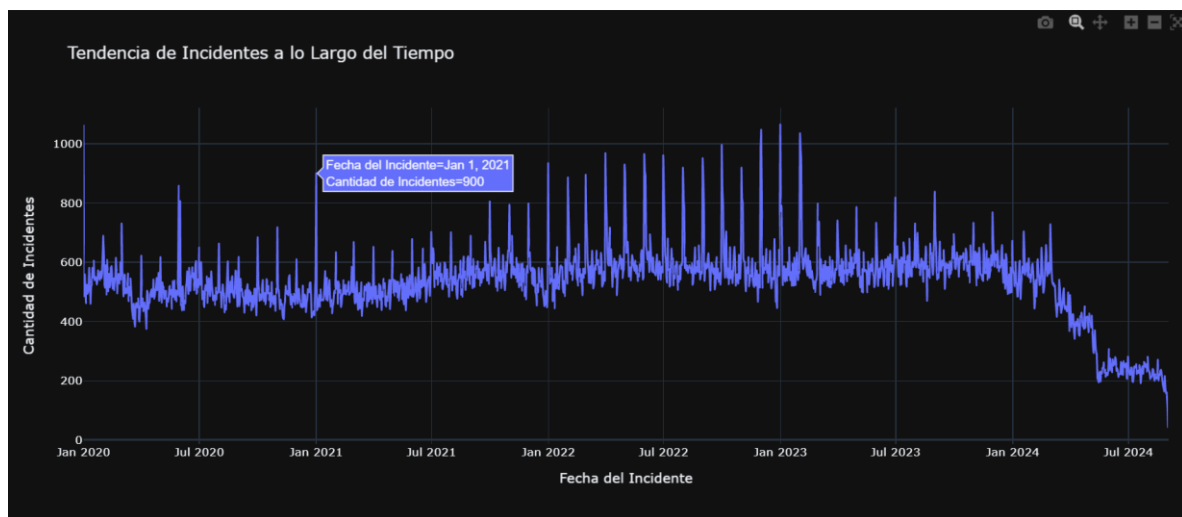
5.1 Tendencia de Incidentes a lo Largo del Tiempo

- **Título:** Evolución Temporal de los Incidentes
- **Descripción:** Examina el número de incidentes reportados a lo largo del tiempo.

Este gráfico muestra un aumento en los reportes durante los primeros días de cada mes, especialmente los días 1 y 2, al igual que se muestran picos en festividades como año nuevo.

También muestra como fueron más constantes los picos entre 2022 y 2023.

Pero es mas notorio el hecho de que a partir de inicios de marzo hubo una notable disminución en el número de incidentes reportados, esto tal vez porque el jefe Choi se convirtió en líder del departamento más grande del país.

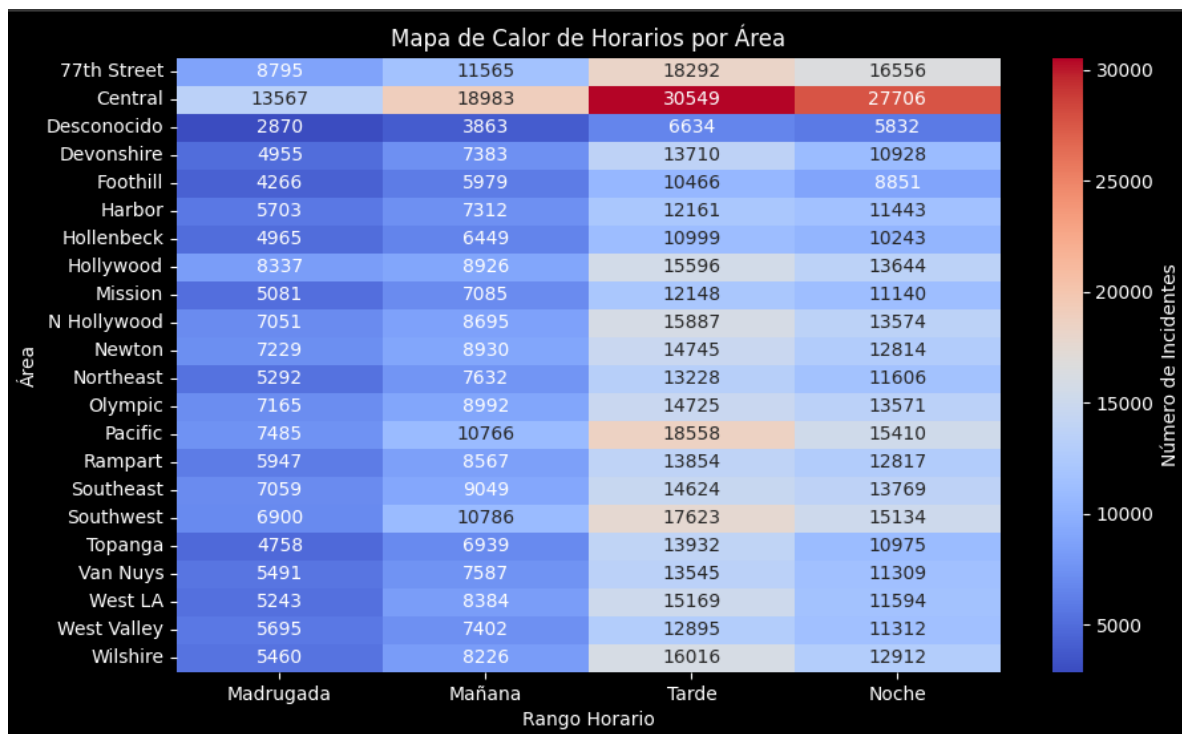


5.2 Mapa de Calor de Horarios por Área

- **Título:** Patrones Horarios en Diferentes Áreas
- **Descripción:** Muestra las horas pico de incidentes según la ubicación.

Los incidentes son más frecuentes durante la **Tarde** en áreas como Central, mientras que otras zonas muestran actividad en horarios diferentes. Esto puede informar estrategias de vigilancia.

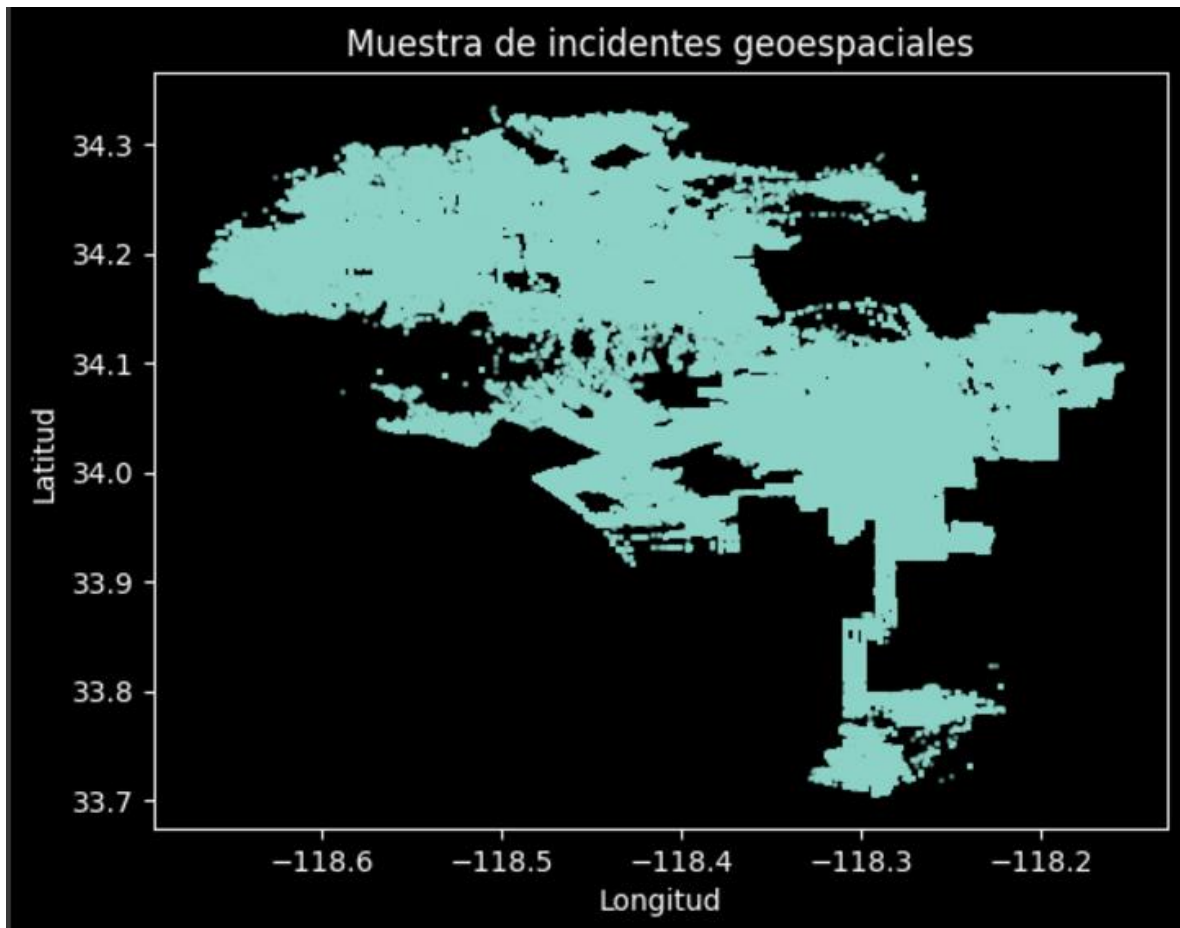
También, en casi todos lados se reportan mas incidentes durante la **Tarde**, seguido de la **Noche**, después en la **Mañana** y por último en la **Madrugada**.



5.3 Mapa de Incidentes Geoespaciales

- **Título:** Visualización Geográfica de Incidentes
- **Descripción:** Mapa que muestra los puntos de los incidentes reportados

Este mapa destaca la dispersión geográfica de los incidentes, mostrando concentraciones claras de cada zona de los Ángeles. La herramienta interactiva facilita identificar patrones específicos en ubicaciones críticas.



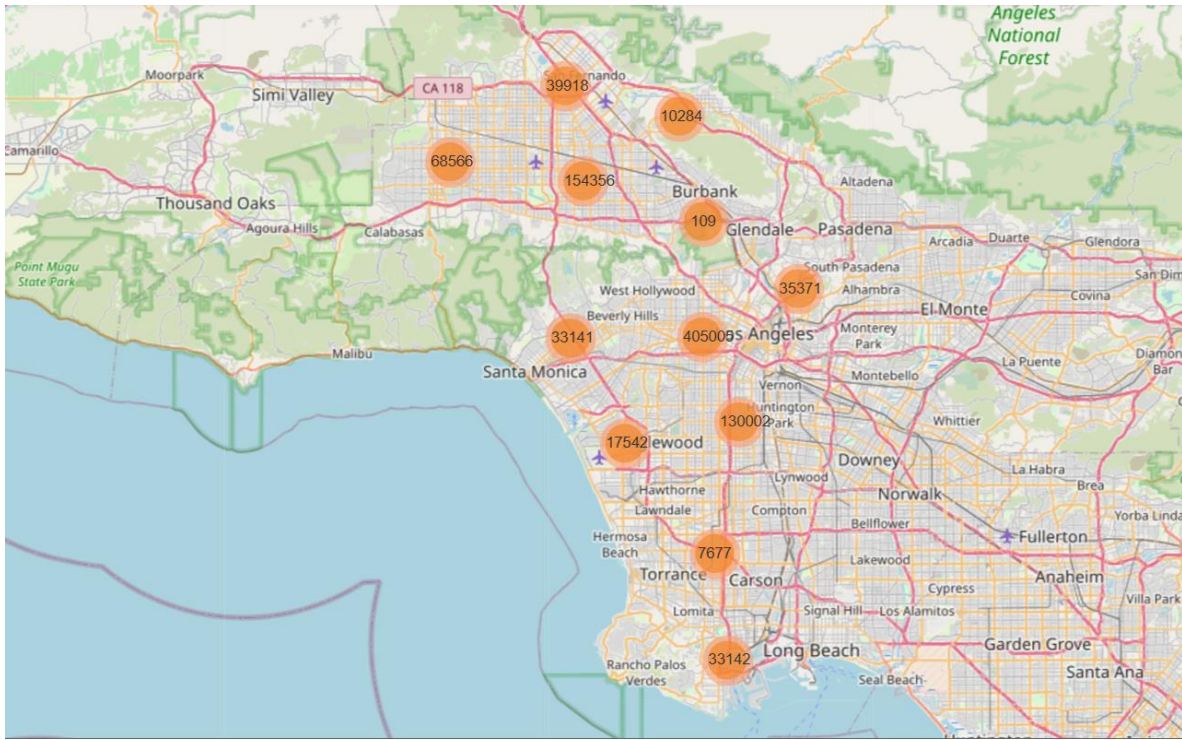
5.4 Mapa Interactivo de Incidentes Geoespaciales

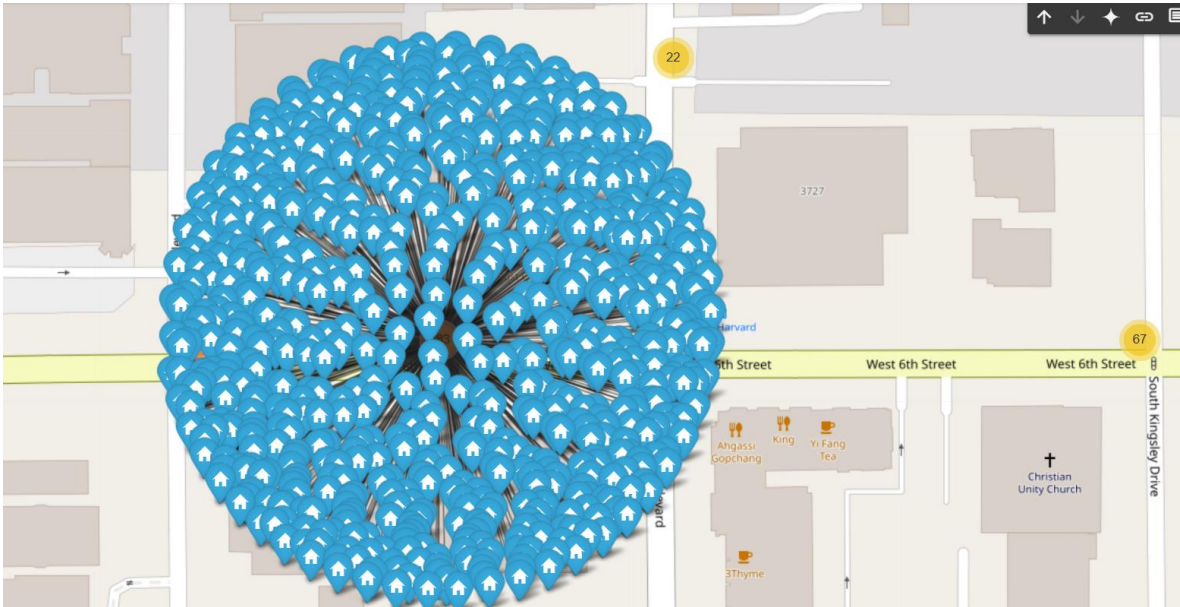
- **Título:** Visualización Geográfica de Incidentes en cada región
- **Descripción:** Mapa que muestra los puntos de los incidentes reportados en cada zona

Se muestra en este mapa interactivo cada zona o calle en la que se han reportado incidentes. La herramienta interactiva facilita identificar patrones específicos en ubicaciones críticas.

Notamos la presencia de que llega a haber hasta más de 600 casos en un solo punto, siendo muy alto para ser un solo lugar y que sea en un tiempo tan breve. Además de que a la esquina siguiente se pueden reportar otros 20, 50 o hasta 100 casos más, siendo concentraciones muy altas en un pequeño espacio.

También se observa que muchos de los incidentes se reportan en las esquinas de las calles.





Conclusiones y Futuras Líneas de Trabajo

Conclusiones:

1. Distribución y patrones de delitos:

- Los delitos más reportados son **robos de vehículos, agresiones simples y robos en vehículos**, lo que refleja la vulnerabilidad de los espacios públicos y la necesidad de medidas específicas para proteger propiedades móviles.
- Los picos en la actividad delictiva ocurren al inicio de cada mes y durante festividades importantes, como Año Nuevo, lo que podría estar relacionado con patrones sociales o económicos.
- La disminución en los incidentes desde marzo de 2024 puede estar vinculada a cambios estratégicos en el liderazgo del LAPD, lo que sugiere que las políticas recientes tienen un impacto positivo.

2. Características de las víctimas:

- Los hombres representan el 55.7% de las víctimas, mientras que las mujeres y géneros no especificados completan el resto. Las edades entre 26 y 35 años son las más afectadas, posiblemente debido a una mayor exposición en la vida laboral y social activa.

- Las víctimas de descendencias hispano/latinas y blancas son las más reportadas, lo que podría estar relacionado con factores demográficos y socioeconómicos específicos de la población de Los Ángeles.

3. Estado de los casos:

- Más del 75% de los casos permanecen en investigación, lo que resalta problemas de saturación o lentitud en el sistema judicial y policial.
- No se identificaron diferencias significativas en el avance de los casos según el género de las víctimas, lo que indica un trato homogéneo en los procesos investigativos.

4. Patrones espacio-temporales:

- Las áreas más peligrosas son **Central**, **77th Street** y **Pacific**, con alta densidad de incidentes reportados. Estas zonas requieren un enfoque prioritario en políticas de vigilancia y prevención.
- Los incidentes son más frecuentes durante las tardes y noches, lo que puede guiar la asignación de recursos policiales según horarios específicos.

Futuras Líneas de Trabajo

1. Mejoras en los Datos:

- **Enriquecimiento de datos:** Incorporar datos adicionales como densidad poblacional, niveles socioeconómicos y registros de intervenciones policiales para contextualizar mejor los incidentes.
- **Resolución geoespacial:** Refinar el nivel de detalle en las ubicaciones para análisis más precisos, utilizando herramientas avanzadas como GIS.

2. Avances Analíticos:

- **Modelos predictivos:** Desarrollar modelos de predicción para identificar zonas y horarios de mayor riesgo, empleando técnicas como árboles de decisión o redes neuronales.
- **Análisis causal:** Investigar el impacto de cambios en políticas de seguridad en la reducción del crimen, mediante análisis de series temporales o estudios de intervención.

3. Visualizaciones Interactivas:

- Ampliar el uso de mapas geoespaciales interactivos para identificar "puntos calientes" y planificar operativos en áreas críticas.

4. Colaboración con Stakeholders:

- Integrar los hallazgos con iniciativas comunitarias y programas de prevención del delito para fortalecer la relación entre la comunidad y el LAPD.
- Presentar los resultados al gobierno local y organizaciones de justicia social para diseñar políticas informadas.

5. Investigaciones Futuras:

- Evaluar las posibles relaciones entre factores climáticos, eventos deportivos o grandes reuniones públicas y la incidencia delictiva.
- Explorar cómo la percepción de seguridad en diferentes comunidades afecta la disposición a reportar delitos.

Con estas mejoras e investigaciones adicionales, el proyecto puede evolucionar hacia una herramienta integral para la prevención y reducción del crimen en Los Ángeles. Esto cumple con el objetivo principal de proporcionar información accionable para optimizar recursos y diseñar estrategias de seguridad efectivas.