



Chicago Crime Análisis de Datos

Integrantes:

**Franklin Canaza Ccori
Hayde Luzmila Humpire Cutipa
Jair Francesco Huaman Canqui
Jhoel Salomon Tapara Quispe**

**Universidad Nacional de San Agustín de Arequipa
Mayo de 2023**

Índice

1. Descripción del Problema	3
2. Descripción de la Base de Datos	3
2.1. Definición de cada Atributo	3
2.2. Tipos de datos	4
2.3. Valores nulos y valores no nulos	5
2.4. Rango de valores	7
2.5. Valores únicos	7

1. Descripción del Problema

La delincuencia es uno de los problemas sociales más importantes de un país, afectando en los ámbitos de la seguridad pública, la sociedad y la economía. Comprender qué factores provocan un aumento de la delincuencia es fundamental para los responsables políticos en sus esfuerzos por reducir la delincuencia y aumentar la calidad de vida de los ciudadanos. Abordamos este problema con el análisis de la base de datos denominada "Chicago Crime", para estimar las tasas de criminalidad en diferentes regiones de la ciudad de Chicago, con la finalidad de que estos datos pueden brindar nuevas perspectivas para comprender el crimen.

2. Descripción de la Base de Datos

La base de datos contiene información de incidentes delictivos que ocurrieron en la ciudad de Chicago desde el 2001. Los datos se extraen del sistema CLEAR (Citizen Law Enforcement Analysis and Reporting) (Análisis e informes de aplicación de la ley ciudadana) del Departamento de Policía de Chicago. Muestra direcciones solo a nivel de bloques para proteger la privacidad de las personas. La base de datos no es íntegra ni precisa debido a que puede contener errores humanos y algunas clasificaciones de delitos pueden estar aun sin verificar.

2.1. Definición de cada Atributo

La base de datos contiene 22 atributos, la definición de cada uno de los atributos es la siguiente:

1. **unique_key:** Identificador único para el registro.
2. **case_number:** El Número RD (Número de División de Registros) del Departamento de Policía de Chicago, que es único para el incidente.
3. **date:** Fecha en que ocurrió el incidente, esto es a veces una estimación.
4. **block:** La dirección parcialmente redactada donde ocurrió el incidente, colocándola en el mismo bloque que la dirección real.
5. **iucr:** El código de Reporte de Delitos de Illinois Uniform. Esto está directamente relacionado con el Tipo primario y la Descripción.
6. **primary_type:** La descripción principal del código IUCR.
7. **description:** La descripción secundaria del código IUCR, una subcategoría de la descripción primaria.

8. **location_description:** Descripción del lugar donde ocurrió el incidente.
9. **arrest:** Indica si se realizó un arresto.
10. **domestic:** Indica si el incidente estuvo relacionado con el hogar según lo define la Ley de Violencia Doméstica de Illinois.
11. **beat:** Indica el área geográfica.
12. **district:** Indica el distrito policial donde ocurrió el incidente.
13. **ward:** El barrio (distrito del Ayuntamiento) donde ocurrió el incidente.
14. **community_area:** Indica el área de la comunidad donde ocurrió el incidente. Chicago tiene 77 áreas comunitarias.
15. **fbi_code:** Indica la clasificación del delito como se describe en el Sistema Nacional de Informes Basados en Incidentes (NIBRS).
16. **x_coordinate:** La coordenada x de la ubicación donde ocurrió el incidente en la proyección State Plane Illinois East NAD 1983.
17. **y_coordinate:** La coordenada y del lugar donde ocurrió el incidente en la proyección State Plane Illinois East NAD 1983.
18. **year:** Año en que ocurrió el incidente.
19. **updated_on:** Fecha y hora en que se actualizó el registro por última vez.
20. **latitude:** La latitud del lugar donde ocurrió el incidente.
21. **longitude:** La longitud del lugar donde ocurrió el incidente.
22. **location:** La ubicación donde ocurrió el incidente en un formato que permite la creación de mapas y otras operaciones geográficas.

2.2. Tipos de datos

Los tipos de datos asociados a cada atributo son:

Tabla de tipos de datos	
unique_key	INTEGER
case_number	STRING
date	TIMESTAMP
block	STRING
iucr	STRING
primary_type	STRING
description	STRING
location_description	STRING
arrest	BOOLEAN
domestic	BOOLEAN
beat	INTEGER
district	INTEGER
ward	INTEGER
community_area	INTEGER
fbi_code	STRING
x_coordinate	FLOAT
y_coordinate	FLOAT
year	INTEGER
updated_on	TIMESTAMP
latitude	FLOAT
longitude	FLOAT
location	STRING

En los datos podemos observar que existen atributos con el mismo tipo, en la siguiente tabla mostramos el número de atributos que pertenecen a cada tipo de dato:

Número de atributos por tipo	
INTEGER	6
STRING	8
TIMESTAMP	2
BOOLEAN	2
FLOAT	4
Total	22

2.3. Valores nulos y valores no nulos

En la siguiente tabla mostramos los atributos que aceptan valores nulos y no nulos:

Tabla de tipos de datos			
unique_key	REQUIRED	case_number	NULLABLE
date	NULLABLE	block	NULLABLE
iucr	NULLABLE	primary_type	NULLABLE
description	NULLABLE	location_description	NULLABLE
arrest	NULLABLE	domestic	NULLABLE
beat	NULLABLE	district	NULLABLE
ward	NULLABLE	community_area	NULLABLE
fbi_code	NULLABLE	x_coordinate	NULLABLE
y_coordinate	NULLABLE	year	NULLABLE
updated_on	NULLABLE	latitude	NULLABLE
longitude	NULLABLE	location	NULLABLE

Iniciamos el análisis con datos filtrados del año 2017 conteniendo en total 268806 registros, el numero total de valores nulos y no nulos por atributo se describen en la siguiente tabla:

Valores nulos y no nulos		
Atributo	No nulos	Nulos
unique_key	268806	0
case_number	268806	0
date	268806	0
block	268806	0
iucr	268806	0
primary_type	268806	0
description	268806	0
location_description	268806	0
arrest	268806	0
domestic	268806	0
beat	268806	0
district	268806	0
ward	268806	0
community_area	268806	0
fbi_code	268806	0
x_coordinate	264934	3872
y_coordinate	264934	3872
year	268806	0
updated_on	268806	0
latitude	264934	3872
longitude	264934	3872
location	264934	3872

2.4. Rango de valores

Dentro de la base de datos podemos identificar intervalos temporales(date, update on y year), intervalos espaciales(x coordinate, y coordiante, latitude y longitude) e intervalos para el identificador (unique key).

Rango de valores		
Atributo	Mínimo	Máximo
date, update_on	01/01/2001	31/12/2020
unique_key	23 059	12 378 294
x_coordinate	1 094231.0	1 205119.0
y_coordinate	1 813 909	1 951 535.0
year	2017	2017
latitude	41.644606	42.022671
longitude	-87.928909	-87.524529

2.5. Valores únicos

Según el portal web de Chicago, se tiene 100 codigos unicos para los Reportes de Delitos de Illinois Unifrom(IUCR), junto con los atributos de primary type y description(secondary type).

IUCR	PRIMARY DESCRIPTION	SECONDARY DESCRIPTION
110	HOMICIDE	FIRST DEGREE MURDER
261	CRIMINAL SEXUAL ASSAULT	AGGRAVATED - HANDGUN
312	ROBBERY	ARMED - KNIFE / CUTTING INSTRUMENT
041A	BATTERY	AGGRAVATED - HANDGUN
470	PUBLIC PEACE VIOLATION	RECKLESS CONDUCT
475	BATTERY	OF AN UNBORN CHILD
490	RITUALISM	AGGRAVATED RITUAL MUTILATION - HANDGUN
495	BATTERY	AGGRAVATED OF A SENIOR CITIZEN
510	RITUALISM	AGG. RITUAL MUTILATION - HANDS
051A	ASSAULT	AGGRAVATED - HANDGUN
580	STALKING	SIMPLE
610	BURGLARY	FORCIBLE ENTRY
810	THEFT	OVER \$500