

Bitacora 2

Bitácora 2 Proyecto Grupal

1

Características de los datos

Se utilizaron las bases de datos de crímenes registrados entre 2019 y julio de 2025, disponibles en el portal de datos del Poder Judicial: <https://sitiooij.poder-judicial.go.cr/index.php/ayuda/servicios-policiales/servicios-a-organizaciones/indice-de-transparencia-del-sector-publico-costarricense/datos-abiertos>.

En este sentido, estas bases venían separadas por año y presentaban algunas inconsistencias entre ellas. Por ejemplo, entre 2019 y 2023 la variable del sexo se denominaba “género”, mientras que a partir de 2024 pasó a llamarse “sexo”. Asimismo, la columna de fecha aparecía en formato *char* en algunos años y en formato *Date* en otros.

Para resolver esto, se unificaron los datos, combinando todos los archivos en una única base que incluye todos los crímenes registrados en el periodo 2019–julio 2025.

De esta forma, resultó una base de datos con las siguientes características:

- Tamaño: 345065 x 12.
- Va de 2019-01-01 a 2025-07-31.
- Todas las variables son cualitativas: delito, subdelito, fecha, hora, victim, subvictim, edad, sexo, nacionalidad, provincia, canton, distrito.

Población de estudio

La población de estudio corresponde a todos los eventos delictivos ocurridos en Costa Rica de 2019 a julio de 2025.

Muestra observada

La muestra son las observaciones registradas de delitos en Costa Rica del 01/01/2019 a las 00:00:00 al 31/07/25 a las 23:59:59, almacenadas en la base del Poder Judicial.

Unidad estadística

Corresponde a cada observación de delito junto con sus atributos: delito, subdelito, fecha, hora, victim, subvictim, edad, sexo, nacionalidad, provincia, canton, distrito

Variables de estudio

Variable	Tipo	Descripción
fecha	Fecha	Día del evento
hora	Categórica	Tramos horarios de 3 horas
delito	Categórica	Categoría principal del hecho
subdelito	Categórica	Subtipo de delito
victima	Categórica	Tipo de víctima (pueden ser objetos como VIVIENDA o VEHÍCULO).
subvictim	Categórica	Subtipo de víctima
sexo	Categórica	Sexo de la víctima.
edad	Categórica	Edad de la víctima.
nacionalidad	Categórica	Nacionalidad de la víctima.
provincia	Categórica	Provincia donde ocurrió el hecho
canton	Categórica	Cantón donde ocurrió el hecho
distrito	Categórica	Distrito donde ocurrió el hecho

Es importante recalcar que las bases de datos traen columnas correspondientes a persona (edad, sexo, nacionalidad) incluso cuando victim es VEHICULO, VIVIENDA, EDIFICACION, etc; lo cual se puede deber a valores por defecto de captura. Para un correcto análisis se coloca NA en dichas posiciones. hacer

2

```
slice_head(datos, n = 5)
```

```
# A tibble: 5 x 12
  delito subdelito fecha      hora  victimam subvictima edad   sexo nacionalidad
  <chr>  <chr>     <date>    <chr> <chr>    <chr> <chr> <chr> <chr>
1 ASALTO ARMA BLAN~ 2019-12-31 21:0~ PERSONA PEATON [P~ Mayo~ HOMB~ COSTA RICA
2 ASALTO ARMA BLAN~ 2019-12-31 00:0~ PERSONA PEATON [P~ Mayo~ HOMB~ COSTA RICA
3 ASALTO ARMA BLAN~ 2019-12-31 00:0~ PERSONA OTRO O IN~ Mayo~ HOMB~ COSTA RICA
4 ASALTO ARMA BLAN~ 2019-12-31 18:0~ PERSONA PEATON [P~ Mayo~ HOMB~ COSTA RICA
5 ASALTO ARMA CONT~ 2019-12-31 21:0~ PERSONA PEATON [P~ Mayo~ HOMB~ NICARAGUA
# i 3 more variables: provincia <chr>, canton <chr>, distrito <chr>
```

3

La base de datos no presenta variables cuantitativas; incluso la edad es presentada en categorías (MAYOR DE EDAD o MENOR DE EDAD), por lo tanto no se puede realizar el resumen.