Ciudad de México, a 29 de enero de 2020

Nota técnica
Asunto: Desglose, desagregación e interpretación de los recursos del conjunto de datos “Hechos de tránsito registrados por la Secretaría de Seguridad Ciudadana”

A. CONTEXTO

A partir de enero de 2024, el conjunto de datos **“Hechos de tránsito registrados por la SSC (serie completa)”** fue reestructurado con el fin de facilitar la presentación de la información contenida en la tabla de datos original elaborada por la Secretaría de Seguridad Ciudadana.

Derivado de esta reestructuración, el conjunto de datos se dividió en cinco recursos que presentan la información con diferente nivel de desagregación. El objetivo de este documento es facilitar la interpretación de la información de cada uno de estos recursos así como documentar el proceso que se realiza a partir de los datos fuente enviados por la Secretaría de Seguridad Ciudadana.

B. CARACTERÍSTICAS DE LA TABLA DE DATOS FUENTE DE LA SSC

En primer lugar, es necesario especificar la estructura de la tabla de datos fuente elaborada por la Secretaría de Seguridad Ciudadana. Esta tabla de datos tiene las siguientes características:

1. La unidad de observación es el hecho de tránsito

Esto significa que cada fila de la tabla representa un hecho de tránsito y toda la información relativa a ese hecho de tránsito viene determinada en cada una de las columnas de la tabla.

Ejemplo ilustrativo

FOLIO	FECHA EVENTO	HORA EVENTO	TIPO EVENTO	COORDENADA X	COORDENADA Y	PUNTO 1
C5/20230101/85967	01/01/2023	09:36	CHOQUE	19.367860	-98.998560	CALZ GRAL IGNACIO ZARAGOZA

2. La información que caracteriza a los hechos de tránsito viene distribuída en 24 columnas:

```
1. fecha_evento
2. hora_evento
3. tipo_evento
4. fecha_captura
5. folio
6. latitud
7. longitud
8. punto_1
9. punto_2
10.colonia
11.alcaldia
12.zona_vial
13.sector
14.unidad_a_cargo
15.tipo_de_interseccion
16.interseccion_semaforizada
17.clasificacion_de_la_vialidad
18.sentido_de_circulacion
19.dia
20.prioridad
21.origen
22.unidad_medica_de_apoyo
23.matricula_unidad_medica
24.trasladado_lesionados
```

3. **La información que caracteriza a las personas lesionadas y fallecidas viene distribuida en 19 columnas**

```
1. total_occisos
2. total_lesionados
3. occisos_femeninos
4. occisos_masculinos
5. occisos_se_desconoce
6. lesionados_femeninos
7. lesionados_masculinos
8. lesionados_se_desconoce
9. edad_occiso
10.edad_lesionado
11.peaton_occiso
12.peaton_lesionado
13.ciclista_occiso
14.ciclista_lesionado
15.pasajero_occiso
16.pasajero_lesionado
17.motociclista_occiso
18.motociclista_lesionado
19.conductor_occiso
```

Estas 19 columnas son de tipo numérico e indican la cantidad de personas registradas en cada uno de estos rubros.

4. **La información que caracteriza a los vehículos involucrados viene distribuida en 3 o 4 columnas dependiendo del número de vehículos involucrados en un hecho de tránsito**

```
1. tipo_de_vehiculo_1
2. tipo_de_vehiculo_2
3. tipo_de_vehiculo_3
4. tipo_de_vehiculo_4
```

5. **La tabla de datos tiene un formato conocido como *wide* o "ancho".**

Los formatos de datos anchos se caracterizan por distribuir la información con una mayor cantidad de columnas en lugar de distribuirla en filas. En el caso de la tabla de datos fuente, esto puede dificultar la interpretación de la información debido a que contiene información de personas (lesionados o fallecidos) y de vehículos, pero conserva como unidad de observación el hecho de tránsito en lugar de la persona o el vehículo.

A continuación se muestra la forma en la que se organizó y separó la información para presentarla en un formato *long* o "largo" para mejorar la interpretación de los datos.

C. RECURSOS REESTRUCTURADOS Y SU CORRECTA INTERPRETACIÓN

I. Hechos de tránsito

Descripción: Este recurso contiene la información de las 24 columnas de caracterización de los hechos de tránsito y dos columnas adicionales:

1. ***total_occisos*** renombrada ***personas_fallecidas***
2. ***total_lesionados*** renombrada ***personas_lesionadas***

Unidad de observación: La unidad de observación de esta tabla de datos es el hecho de tránsito. En este sentido cada fila representa un hecho de tránsito.

Interpretación: Esta tabla de datos es útil para realizar un conteo de los hechos de tránsito a través del tiempo y desagregarse por tipo de evento, alcaldía o cualquier otra variable de interés. Además, permite hacer un conteo simple del total de personas lesionadas y fallecidas.

II. Personas lesionadas o fallecidas (por sexo)

Descripción: Este recurso contiene la información de las 24 columnas de caracterización de los hechos de tránsito y tres columnas adicionales:

1. ***sexo***: con tres valores: *"Hombre", "Mujer", "Desconocido"*.
2. ***condicion_persona:*** con dos valores: *"lesionada", "fallecida"*
3. ***total:*** con el total numérico de personas registradas en cada categoría

Unidad de observación: La unidad de observación de esta tabla de datos es la relación sexo/condición persona. En este sentido cada hecho de tránsito puede repetirse hasta en 6 filas dependiendo de la cantidad de personas lesionadas o fallecidas para cada categoría de la columna "sexo"

Interpretación: Esta tabla de datos tiene como objetivo facilitar el conteo de personas lesionadas y fallecidas en hechos de tránsito desglosados por sexo.

Ejemplo ilustrativo de la tabla original

Tabla de datos fuente SSC

FOLIO	OCCISOS FEMENINOS	OCCISOS MASCULINOS	OCCISOS SE DESCONOCE	LESIONADOS FEMENINOS	LESIONADOS MASCULINOS	LESIONADOS SE DESCONOCE
C5/20230101/85967	0	1	0	1	1	2

Ejemplo ilustrativo de la tabla reestructurada con la misma información

Tabla reestructurada de datos abiertos

folio	sexo	condicion_persona	total
C5/20230101/85967	Hombre	Lesionada	1
C5/20230101/85967	Mujer	Lesionada	1
C5/20230101/85967	Desconocido	Lesionada	2
C5/20230101/85967	Hombre	Fallecida	1

III. Personas lesionadas o fallecidas (por edad)

Descripción: Este recurso contiene la información de las 24 columnas de caracterización de los hechos de tránsito y dos columnas adicionales:

1. ***condicion_persona:*** con dos valores: *"lesionada", "fallecida"*
2. ***edad:*** con la edad de la persona

Unidad de observación: La unidad de observación de esta tabla de datos es la persona. En este sentido cada hecho de tránsito puede repetirse hasta en *"n"* cantidad de filas dependiendo de la cantidad de personas lesionadas o fallecidas.

Interpretación: El objetivo de este recurso es facilitar el análisis de edades por persona y los agregados como promedio de edad.

Aviso: Es posible que el número de personas lesionadas o fallecidas en esta tabla de datos discrepe del total en otras tablas si la edad de la persona lesionada o fallecida no se registró.

Ejemplo ilustrativo de la tabla original

Tabla de datos fuente SSC

FOLIO	EDAD OCCISO	EDAD LESIONADO
C5/20230101/85967	NA	15 25 35 35

Ejemplo ilustrativo de la tabla reestructurada con la misma información

Tabla reestructurada de datos abiertos

folio	condicion_persona	edad
C5/20230101/85967	Lesionada	15
C5/20230101/85967	Lesionada	25
C5/20230101/85967	Lesionada	35
C5/20230101/85967	Lesionada	35

IV. Personas lesionadas o fallecidas (por tipo de persona usuaria de vía)

Descripción: Este recurso contiene la información de las 24 columnas de caracterización de los hechos de tránsito y tres columnas adicionales:

1. ***tipo_persona:*** con cinco valores: *"Peatón", "Ciclista", "Pasajera", "Conductora", "Motociclista".*
2. ***condicion_persona:*** con dos valores: *"lesionada", "fallecida"*
3. ***total:*** con el total numérico de personas registradas en cada categoría

Unidad de observación: La unidad de observación de esta tabla de datos es la relación tipo de persona/condición persona. En este sentido cada hecho de tránsito puede repetirse hasta en 10 filas dependiendo de la cantidad de personas lesionadas o fallecidas para cada categoría de la columna "tipo_persona".

Interpretación: Esta tabla de datos tiene como objetivo facilitar el conteo de personas lesionadas y fallecidas en hechos de tránsito desglosados por tipo de persona usuaria de vía.

Ejemplo ilustrativo de la tabla original

Tabla de datos fuente SSC

FOLIO	PEATON OCCISO	PEATON LESIONADO	CICLISTA OCCISO	CICLISTA LESIONADO	PASAJERO OCCISO	PASAJERO LESIONADO	MOTOCICLISTA OCCISO	MOTOCICLISTA LESIONADO	CONDUCTOR OCCISO	CONDUCTOR LESIONADO
C5/20230101/85967	0	5	0	1	0	3	1	1	0	1

Ejemplo ilustrativo de la tabla reestructurada con la misma información

Tabla reestructurada de datos abiertos

folio	tipo_persona	condicion_persona	total
C5/20230101/85967	Peatón	Lesionada	5
C5/20230101/85967	Ciclista	Lesionada	1
C5/20230101/85967	Pasajero	Lesionada	3
C5/20230101/85967	Motociclista	Fallecida	1
C5/20230101/85967	Motociclista	Lesionada	1
C5/20230101/85967	Conductor	Lesionada	1

V. Vehículos involucrados en hechos de tránsito.

Descripción: Este recurso contiene la información de las 24 columnas de caracterización de los hechos de tránsito y dos columnas adicionales:

1. ***no_vehiculo:*** con el identificador del vehículo involucrado en el hecho de tránsito: *"Vehículo 1", "Vehículo 2", … "Vehículo n".*
2. ***tipo_vehiculo:*** contiene un catálogo de 17 valores posibles. Este catálogo puede expandirse de acuerdo con los reportes de cada actualización de datos.

Unidad de observación: La unidad de observación de esta tabla de datos es el vehículo. En este sentido cada hecho de tránsito puede repetirse hasta en *"n"* cantidad de filas dependiendo de la cantidad de vehículos involucrados.

Interpretación: El objetivo de este recurso es facilitar el análisis del conteo de vehículos involucrados en hechos de tránsito desglosados por tipo de vehículo.

Ejemplo ilustrativo de la tabla original

Tabla de datos fuente SSC

FOLIO	TIPO DE VEHÍCULO 1	TIPO DE VEHÍCULO 2	TIPO DE VEHÍCULO 3
C5/20230101/85967	AUTOMOVIL	BICICLETA	CAMIONETA

Ejemplo ilustrativo de la tabla reestructurada con la misma información

Tabla reestructurada de datos abiertos

folio	no_vehiculo	tipo_vehiculo
C5/20230101/85967	Vehículo 1	AUTOMOVIL
C5/20230101/85967	Vehículo 2	BICICLETA
C5/20230101/85967	Vehículo 3	CAMIONETA

D. POSIBLES ERRORES DE INTERPRETACIÓN

La reestructuración de la tabla de datos fuente de la Secretaría de Seguridad Ciudadana se realizó conservando el folio y los datos de caracterización de los hechos de tránsito en todos los recursos derivados. De esta manera, la separación y desglose de los datos permite estudiar y analizar toda la información contenida en la base original de la SSC. La naturaleza de la base de datos original, sin embargo, no permite realizar algunos análisis. A continuación se da una lista de las conclusiones que pueden y no pueden extraerse de los datos.

¿Qué puede analizarse con esta base de datos?

1. Es posible conocer el total de personas fallecidas y lesionadas por hecho de tránsito
2. Es posible conocer el total de personas fallecidas y lesionadas desglosadas por sexo para cada hecho de tránsito.
3. Es posible conocer el total de personas fallecidas y lesionadas desglosadas por tipo de persona usuaria de vía para cada hecho de tránsito.
4. Es posible conocer las edades de las personas lesionadas o fallecidas para cada hecho de tránsito únicamente para las personas cuya edad fue registrada.
5. Es posible conocer el total de vehículos involucrados desglosado por tipo de vehículo para cada hecho de tránsito.

¿Qué *no puede* analizarse con esta base de datos?

1. No pueden realizarse análisis a nivel persona, únicamente a nivel hecho de tránsito.
2. No puede conocerse la caracterización completa de una persona fallecida o lesionada. En otras personas, no puede saberse ***simultáneamente*** el sexo, la edad y el tipo de persona usuaria de vía de una persona en particular.