

Tarea 2

Cárdenas Avila Josué David

Para esta tarea tomamos una base de datos que recopila los accidentes ocurridos en Xalapa.

Paso 1: Descripción de la estructura de los datos

Lo primero interesante que observamos aquí es que contamos con datos float, int y object.

```
[31] datos.dtypes

idMES                                int64
DELEGACION DE TRANSITO                object
FECHA DEL ACCIDENTE                  object
DIA DE LA SEMANA                      object
HORA EN QUE SE RECIBIO LA LLAMADA REPORTANDO EL ACCIDENTE  object
...
HOMBRE MUERTO DE 31-45 ANIOS          float64
MUJER MUERTA DE 46-60 ANIOS           float64
HOMBRE MUERTO DE 46-60 ANIOS          float64
MUJER MUERTA MAS DE 60 ANIOS          float64
HOMBRE MUERTO MAS DE 60 ANIOS         float64
Length: 64, dtype: object
```

Paso 2: Identificación de datos faltantes

Lo siguiente que observamos es una gran cantidad de datos nulos, debido a que seguramente no se hicieron de forma correcta los reportes o se trató de profundizar de más como se verá a continuación.

```
[32] datos.isnull().sum()

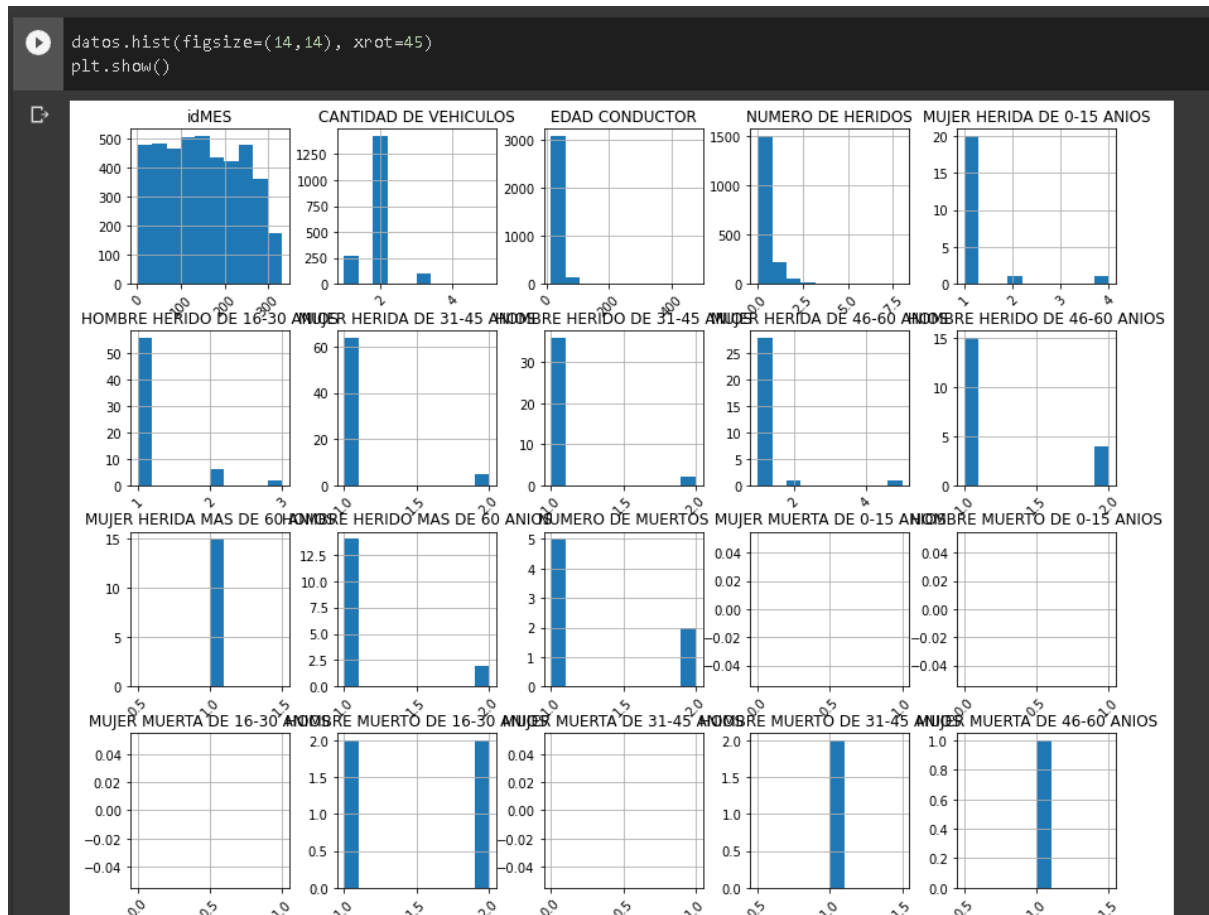
idMES                                0
DELEGACION DE TRANSITO                2506
FECHA DEL ACCIDENTE                  2507
DIA DE LA SEMANA                      2506
HORA EN QUE SE RECIBIO LA LLAMADA REPORTANDO EL ACCIDENTE  3792
...
HOMBRE MUERTO DE 31-45 ANIOS          4309
MUJER MUERTA DE 46-60 ANIOS           4310
HOMBRE MUERTO DE 46-60 ANIOS          4309
MUJER MUERTA MAS DE 60 ANIOS          4311
HOMBRE MUERTO MAS DE 60 ANIOS         4310
Length: 64, dtype: int64
```

Tarea 2

Cárdenas Avila Josué David

Paso 3: Detección de valores atípicos

Como es de esperarse contamos con múltiples datos atípicos debido a la naturaleza de los eventos.



Cárdenas Avila Josué David

4) Distribución de variables categóricas

Finalmente debido a que al menos un campo de cada renglón contaba con falta de información obtenemos la siguiente matriz de correlación, está únicamente nos indica que las variables a las que más deberíamos de prestar atención son edad de conductor y número de muertes.

