

Universidad técnica particular de Loja.

Nombre: Roberto Bustamante

Fecha: 25/04/2025

A. Examinar los datos (estructura, dimensionalidad, etc)

Print.

```
58          EMBRIAGUEZ O DROGA          1          4          5
59 IMPERICIA E IMPRUDENCIA DEL CONDUCTOR 1          1          2
60          EMBRIAGUEZ O DROGA          0          1          1
61 IMPERICIA E IMPRUDENCIA DEL CONDUCTOR 0          0          0
62          EXCESO VELOCIDAD            0          0          0
63 IMPERICIA E IMPRUDENCIA DEL CONDUCTOR 0          1          1
64          EXCESO VELOCIDAD            1          1          2
65 IMPERICIA E IMPRUDENCIA DEL CONDUCTOR 0          3          3
66          NO RESPETA LAS SEÑALES DE TRÁNSITO 0          6          6
67          EXCESO VELOCIDAD            0          0          0
68          EXCESO VELOCIDAD            0          1          1
69          EMBRIAGUEZ O DROGA          0          1          1
70 IMPERICIA E IMPRUDENCIA DEL CONDUCTOR 0          0          0
71 IMPERICIA E IMPRUDENCIA DEL CONDUCTOR 0          1          1
72 IMPERICIA E IMPRUDENCIA DEL CONDUCTOR 0          0          0
73          MAL REBASAMIENTO INVADIR CARRIL 0          0          0
74 IMPERICIA E IMPRUDENCIA DEL CONDUCTOR 0          0          0
75 IMPERICIA E IMPRUDENCIA DEL CONDUCTOR 0          1          1
76          OTRAS CAUSAS                 0          0          0
77          NO RESPETA LAS SEÑALES DE TRÁNSITO 0          0          0
78          EXCESO VELOCIDAD            0          1          1
79          EXCESO VELOCIDAD            0          2          2
80          EMBRIAGUEZ O DROGA          0          0          0
81          NO RESPETA LAS SEÑALES DE TRÁNSITO 0          1          1
82          EXCESO VELOCIDAD            0          2          2
83          EXCESO VELOCIDAD            0          1          1
[ reached 'max' / getOption("max.print") -- omitted 21269 rows ]
```

Head.

```
> head(datossini)
  ANIO MES DIA HORA PROVINCIA CANTÓN ZONA CLASE CAUSA
1 2021 ENERO VIERNES 00:00 A 00:59 TUNGURAHUA AMBATO RURAL CHOQUES EMBRIAGUEZ O DROGA
2 2021 ENERO VIERNES 00:00 A 00:59 CHIMBORAZO RIOBAMBA URBANA CHOQUES IMPERICIA E IMPRUDENCIA DEL CONDUCTOR
3 2021 ENERO VIERNES 00:00 A 00:59 GUAYAS GUAYAQUIL URBANA ATROPELLOS EXCESO VELOCIDAD
4 2021 ENERO VIERNES 00:00 A 00:59 LOS RÍOS PUEBLOVIEJO RURAL ROZAMIENTOS IMPERICIA E IMPRUDENCIA DEL CONDUCTOR
5 2021 ENERO VIERNES 01:00 A 01:59 GUAYAS NARANJAL RURAL CHOQUES NO RESPETA LAS SEÑALES DE TRÁNSITO
6 2021 ENERO VIERNES 01:00 A 01:59 GUAYAS GUAYAQUIL URBANA PÉRDIDA DE PISTA EXCESO VELOCIDAD
  NUM_FALLECIDO NUM_LESIONADO TOTAL_VICTIMAS
1              0              0              0
2              0              1              1
3              0              2              2
4              0              1              1
5              0              2              2
6              0              0              0
```

Str.

```
> str(datossini)
'data.frame': 21352 obs. of 12 variables:
 $ ANIO      : int  2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
 $ MES       : chr  "ENERO" "ENERO" "ENERO" "ENERO" ...
 $ DIA       : chr  "VIERNES" "VIERNES" "VIERNES" "VIERNES" ...
 $ HORA      : chr  "00:00 A 00:59" "00:00 A 00:59" "00:00 A 00:59" ...
 $ PROVINCIA : chr  "TUNGURAHUA" "CHIMBORAZO" "GUAYAS" "LOS RÍOS" ...
 $ CANTÓN    : chr  "AMBATO" "RIOBAMBA" "GUAYAQUIL" "PUEBLOVIEJO" ...
 $ ZONA      : chr  "RURAL" "URBANA" "URBANA" "RURAL" ...
 $ CLASE     : chr  "CHOQUES" "CHOQUES" "ATROPELLOS" "ROZAMIENTOS" ...
 $ CAUSA     : chr  "EMBRIAGUEZ O DROGA" "IMPERICIA E IMPRUDENCIA DEL CONDUCTOR" "EXCESO VELOCIDAD" "IMPERICIA E IMPRUDEN
CIA DEL CONDUCTOR" ...
 $ NUM_FALLECIDO : int  0 0 0 0 0 0 0 0 0 ...
 $ NUM_LESIONADO : int  0 1 2 1 2 0 1 1 1 0 ...
 $ TOTAL_VICTIMAS: int  0 1 2 1 2 0 1 1 1 0 ...
```

Summary.

```
> summary(datossini)
      ANIO      MES      DIA      HORA      PROVINCIA      CANTÓN
Min.   :2021  Length:21352  Length:21352  Length:21352  Length:21352  Length:21352
1st Qu.:2021  Class :character  Class :character  Class :character  Class :character  Class :character
Median :2021  Mode  :character  Mode  :character  Mode  :character  Mode  :character  Mode  :character
Mean   :2021
3rd Qu.:2021
Max.   :2021

      ZONA      CLASE      CAUSA      NUM_FALLECIDO      NUM_LESIONADO      TOTAL_VICTIMAS
Length:21352  Length:21352  Length:21352  Min.   : 0.0000  Min.   : 0.0000  Min.   : 0.0000
Class :character  Class :character  Class :character  1st Qu.: 0.0000  1st Qu.: 0.0000  1st Qu.: 0.0000
Mode  :character  Mode  :character  Mode  :character  Median : 0.0000  Median : 1.0000  Median : 1.0000
Mean   : 0.0998  Mean   : 0.8211  Mean   : 0.9209
3rd Qu.: 0.0000  3rd Qu.: 1.0000  3rd Qu.: 1.0000
Max.   :18.0000  Max.   :29.0000  Max.   :44.0000

> |
```

Dim.

```
> dim(datossini)
[1] 21352  12
> |
```

Class.

```
> class(datossini)
[1] "data.frame"
> |
```

Mode.

```
> mode(datossini)
[1] "list"
> |
```

B. Agregar al menos dos atributos derivados

```
datossini$GRAVEDAD <- ifelse(as.numeric(datossini$NUM_FALLECIDO) > 0,
"Grave", "No Grave")
```

crea una nueva columna llamada gravedad la cual analiza si el numero de muertos es mayor a 0 lo clasifica como grave, caso contrario lo pone como no grave.

	CAUSA	NUM_FALLECIDO	NUM_LESIONADO	TOTAL_VICTIMAS	GRAVEDAD
1	EMBRIAGUEZ O DROGA	0	0	0	No Grave
2	IMPERICIA E IMPRUDENCIA DEL CONDUCTOR	0	1	1	No Grave
3	EXCESO VELOCIDAD	0	2	2	No Grave
4	IMPERICIA E IMPRUDENCIA DEL CONDUCTOR	0	1	1	No Grave
5	NO RESPETA LAS SEÑALES DE TRÁNSITO	0	2	2	No Grave
6	EXCESO VELOCIDAD	0	0	0	No Grave
7	IMPERICIA E IMPRUDENCIA DEL CONDUCTOR	0	1	1	No Grave
8	IMPERICIA E IMPRUDENCIA DEL CONDUCTOR	0	1	1	No Grave
9	IMPERICIA E IMPRUDENCIA DEL CONDUCTOR	0	1	1	No Grave
10	EXCESO VELOCIDAD	0	0	0	No Grave
11	NO RESPETA LAS SEÑALES DE TRÁNSITO	0	2	2	No Grave
12	IMPRUDENCIA DEL PEATÓN	1	0	1	Grave
13	IMPERICIA E IMPRUDENCIA DEL CONDUCTOR	1	1	2	Grave
14	IMPERICIA E IMPRUDENCIA DEL CONDUCTOR	1	1	2	Grave

```
datossini$ES_FIN_SEMANA <- ifelse(datossini$DIA %in% c("SÁBADO",
"DOMINGO"), "Si", "No")
```

```
> head(subset_datos)
```

	ANIO	PROVINCIA	CANTÓN	NUM_FALLECIDO	NUM_LESIONADO	TOTAL_VICTIMAS
126	2021	GUAYAS	EL EMPALME	0	2	2
127	2021	GUAYAS	GUAYAQUIL	0	0	0
128	2021	PICHINCHA	QUITO	0	0	0
129	2021	LOS RÍOS	QUEVEDO	0	1	1
130	2021	SANTO DOMINGO DE LOS	TSÁCHILAS	0	1	1
131	2021	PICHINCHA	QUITO	0	0	0


```
ES_FIN_SEMANA
```

126	Si
127	Si
128	Si
129	Si
130	Si
131	Si

crea una nueva columna llamada `es_fin_semana` la cual se encarga de recopilar si el día en el que ocurrió el siniestro fue un sábado o domingo, con lo cual podemos sacar patrones para analizar cuando es que se producen más siniestros.

C. Extraer un subconjunto de datos (filtrando tanto filas como columnas)

```
subset_datos <- subset(datossini,(TOTAL_VICTIMAS) >= 6,select = c("ANIO",
"PROVINCIA", "CANTÓN", "NUM_FALLECIDO", "NUM_LESIONADO",
"TOTAL_VICTIMAS"))
```

Se encarga de filtrar los datos para que los que se muestren sean los siniestros donde las victimas totales hayan sido mayores o iguales a 6.

```
> head(subset_datos)
```

	ANIO	PROVINCIA	CANTÓN	NUM_FALLECIDO	NUM_LESIONADO	TOTAL_VICTIMAS
66	2021	GUAYAS	SAN JACINTO DE YAGUACHI	0	6	6
145	2021	GUAYAS	GUAYAQUIL	0	8	8
165	2021	ORELLANA	ORELLANA	1	5	6
185	2021	GUAYAS	GUAYAQUIL	0	7	7
186	2021	GUAYAS	GUAYAQUIL	0	7	7
468	2021	PICHINCHA	QUITO	0	7	7

D. Para datos numéricos obtenga estadísticas globales (media, mediana, moda, varianza, desviación típica, rango)

```
> mean(datossini$TOTAL_VICTIMAS)
[1] 0.9208973
> median(datossini$TOTAL_VICTIMAS)
[1] 1
> max(datossini$TOTAL_VICTIMAS)
[1] 44
> min(datossini$TOTAL_VICTIMAS)
[1] 0
> sum(datossini$TOTAL_VICTIMAS)
[1] 19663
> sd(datossini$TOTAL_VICTIMAS)
[1] 1.143316
> var(datossini$TOTAL_VICTIMAS)
[1] 1.30717
> range(datossini$TOTAL_VICTIMAS)
[1] 0 44
> diff(range(datossini$TOTAL_VICTIMAS))
[1] 44
```

E. Obtenga estadísticas mediante agrupamientos (Investigar).

```
install.packages("dplyr")
```

```
library(dplyr)
```

```
datossini %>%
```

```
  group_by(PROVINCIA) %>%
```

```
  summarise(
```

```
    promedio_victimas = mean(as.numeric(TOTAL_VICTIMAS), na.rm = TRUE),
```

```
    mediana_victimas = median(as.numeric(TOTAL_VICTIMAS), na.rm = TRUE),
```

```
    desv_tipica_victimas = sd(as.numeric(TOTAL_VICTIMAS), na.rm = TRUE),
```

```
    accidentes = n()
```

```
)
```

Se instala la librería dplyr y luego se la ejecuta, luego se procede a armar la consulta por agrupamiento de los datos por provincia con la cual obtenemos los siguientes resultados.

```
# A tibble: 24 × 5
  PROVINCIA promedio_victimas mediana_victimas desv_tipica_victimas accidentes
  <chr>          <dbl>          <dbl>          <dbl>          <int>
1 AZUAY          0.871            1            1.08            828
2 BOLÍVAR        2.02            1            1.76            80
3 CARCHI         1.21            1            1.31            73
4 CAÑAR          1.79            1            3.93            62
5 CHIMBORAZO     0.752            0            1.46           622
6 COTOPAXI       1.51            1            1.34           104
7 EL ORO         0.967            1            1.03           783
8 ESMERALDAS     1.19            1            1.29           232
9 GALÁPAGOS      1              1            NA              1
10 GUAYAS        1.07            1            0.961          7758
```

Se usa el `as.numeric` para que transforme los datos que se encuentren en tipo texto a tipo numérico y así poder realizar operaciones sin problemas