



**CENTRO POLITÉCNICO SUPERIOR
MALVINAS ARGENTINAS**

DESCRIPCIÓN DEL DATASET

El nombre del dataset utilizado para este proyecto lleva como título **Accidentes del 2009 al 2023 Municipio de Río Grande.**

La misma contiene información relacionada con la cantidad de accidentes viales ocurridos en nuestra ciudad, el objetivo del dataset es brindar una base sólida para el análisis y modelado de variables asociadas a la accidentología vial, como víctimas, lesionados, zonas de ocurrencia y presencia de alcohol.

La fuente de datos corresponde a la Dirección de Tránsito del Municipio de Río Grande, en formato excel y la cual fue otorgada en mayo del corriente año.

El mencionado contiene registros mensuales de los accidentes ocurridos entre los años 2009 y 2023, donde cada año se desglosa por mes e incluye el número total de hechos, víctimas fatales, lesionados, distribución según la zona de ocurrencia (Urbana o Rural) y número total de alcohol positivo registrados en el mes y año.

El dataset otorgado contiene un aproximado de 180 registros anuales.

Cantidad de filas : 200

cantidad de columnas: 15

Nombre de columnas :

Año: 2009-2023 (Categórico)

Mes: mes de registro de enero a diciembre (Categórico)

Hechos Totales: corresponde al número total de accidentes registrados (Numérico)

Víctimas Fatales: Cantidad de muertes registradas (Numérico)

Lesionados: Cantidad total de personas heridas. (Numérico)

Zona: Clasificación de la zona del hecho (Urbana o Rural) [Categórico]

Alcohol Positivo: Cantidad de conductores con test de alcoholemia positiva
(Numérico).

En el año **2020** no se tomaron registros debido a la emergencia sanitaria por COVID-19



CENTRO POLITÉCNICO SUPERIOR MALVINAS ARGENTINAS

Adjunto print de pantalla con el primer dataset
otorgado por el municipio.

https://docs.google.com/spreadsheets/d/1z0wjQDFPSE-YZluS8HVFJwuyVBjNuyM/edit?usp=drive_link&ouid=115329342559361473046&rtpof=true&sd=true

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Accidentes de tránsito, hechos, víctimas y lesionados y hechos por zona de ocurrencia.										Localidad de Río Grande.				
2	Años 2009-2023														
3	Año	Mes	Total	Hechos simples	mortal		con lesiones		Hechos por zona de ocurrencia		Alcohol Positivo				
4					Hechos	Victimas	Hechos	Lesionados	Urbano	Rural		Total			
5															
6	2009	Total	2.694	2.568	5	5	121	163	2.622	72	487				
7		ene	157	154	-	-	3	3	149	8	25				
8		feb	178	176	-	-	2	2	166	12	43				
9		mar	267	250	2	2	15	15	259	8	35				
10		abr	217	202	1	1	14	13	209	8	36				
11		may	233	225	-	-	8	10	224	9	43				
12		jun	288	279	-	-	9	11	282	6	56				
13		jul	261	254	-	-	7	15	257	4	41				
14		ago	240	234	1	1	5	10	239	1	24				
15		sep	220	203	1	1	16	22	217	3	39				
16		oct	198	186	-	-	12	16	190	8	29				
17		nov	217	202	-	-	15	24	217	-	47				
18		dic	218	203	-	-	15	22	213	5	49				
19	2010	Total	3.333	3.183	4	9	146	208	3.224	109	604				
20		ene	164	156	-	-	8	19	150	14	25				
21		feb	199	193	-	-	6	8	187	12	49				
22		mar	281	267	1	1	13	20	269	12	53				
23		abr	270	254	-	-	16	27	260	10	47				
24		may	265	253	-	-	12	16	260	5	31				

print pasado a CSV

https://docs.google.com/spreadsheets/d/15s7HRB-VhYHPtixImRrFUKbtJOLxhMg89Ue_H9HO2KU/edit?usp=drive_link



Accidentes de tránsito, hechos, víctimas y lesionados y hechos por zona de ocurrencia. Localidad de R.A-o Grande. Años 2009-2023,.....

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Accidentes de tránsito, hechos, víctimas y lesionados y hechos por zona de ocurrencia. Localidad de R.A-o Grande. Años 2009-2023,.....															
2	Año,Mes,.....															
3	,, Total , Hechos simples , mortal ,, con lesiones ,, Hechos por zona de ocurrencia ,,Alcohol Positivo															
4	,,, Hechos , Víctimas , Hechos , Lesionados , Urbano , Rural ,Total															
5	2009, ene , 157 , 154 , - , - , 3 , 3 , 149 , 8,25															
6	2009, feb , 178 , 176 , - , - , 2 , 2 , 166 , 12,43															
7	2009, mar , 267 , 250 , 2 , 2 , 15 , 15 , 259 , 8,35															
8	2009, abr , 217 , 202 , 1 , 1 , 14 , 13 , 209 , 8,36															
9	2009, may , 233 , 225 , - , - , 8 , 10 , 224 , 9,43															
10	2009, jun , 288 , 279 , - , - , 9 , 11 , 282 , 6,56															
11	2009, jul , 261 , 254 , - , - , 7 , 15 , 257 , 4,41															
12	2009, ago , 240 , 234 , 1 , 1 , 5 , 10 , 239 , 1,24															
13	2009, sep , 220 , 203 , 1 , 1 , 16 , 22 , 217 , 3,39															
14	2009, oct , 198 , 186 , - , - , 12 , 16 , 190 , 8,29															
15	2009, nov , 217 , 202 , - , - , 15 , 24 , 217 , - ,47															
16	2009, dic , 218 , 203 , - , - , 15 , 22 , 213 , 5,49															
17	2010, ene , 164 , 156 , - , - , 8 , 19 , 150 , 14,25															
18	2010, feb , 199 , 193 , - , - , 6 , 8 , 187 , 12,49															
19	2010, mar , 281 , 267 , 1 , 1 , 13 , 20 , 269 , 12,53															
20	2010, abr , 270 , 254 , - , - , 16 , 27 , 260 , 10,47															
21	2010, may , 265 , 253 , - , - , 12 , 16 , 260 , 5,31															

Proceso ETL:

para realizar el proceso ETL, subi el scv a visual studio code ,trabaje con colab

- se quitaron los colores de las columnas y filas del dataset original
- se elimino filas que no corresponden.
- los espacios sobrantes en los nombres de las columnas las estandarice por ejemplo: "alcohol positivo" por "alcohol_positivo"
- Reemplace '-' por 0 en todo el DataFrame

Todas las cantidades (accidentes totales, accidentes urbanos, rurales, muertos, lesionados, casos con alcohol, etc.) aparecen como enteros sin decimales ni celdas vacías se obtuvo un dataset homogéneo, con tipos de dato correctos y sin filas inválidas ni valores faltantes.