

Integrantes:

- Manuel Valenzuela
- Luisa Arboleda
- Jennifer Barillas

Informe

Para la limpieza de datos necesitábamos inspeccionar cada uno de los archivos, como formas importantes del mismo están los hechos de tránsito, vehículos involucrados y fallecidos y lesionados. Estas tres partes nos ayudarán a evaluar los diferentes accidentes a lo largo del 2009 - 2017, con esto se evalúa la forma en la que son adjuntos los datos.

❖ Adjunción de datos

- Para esta parte bajamos los archivos de la pagina “<https://www.ine.gob.gt/index.php/estadisticas-continuas/accidentes-de-transito>” por año, algunos años tienen las bases de datos con la extensión .sav por lo que se tuvo que descargar y convertir a un archivo de xlsx para tener todos los archivos con la misma extensión.

❖ Análisis

- En los tres archivos importantes de cada año comparamos las columnas o número de variables, sin embargo nos dimos cuenta que cada año era diferente que solo algunas columnas se repetían en todos. Por lo que necesitamos tener la misma cantidad de columnas para poder unir todo exitosamente. Para esto hicimos unos archivos de excel y con ello examinamos cuáles variables harían falta en cada año.

• Vehículos involucrados

2017	2016	2015	2014	2013	2012	2011	2010	2009		
Numero de correlacion	Numero de correlacion	Numero de correlacion	Numero de correlacion	numero de correlacion	numero	Numero de correlacion	Numero	Numero de correlacion	num_corre	Agregadas
corre base	corre base	corre base	corre base	corre base	corre base	corre base	corre base	corre base	corre_base	
Año ocurrencia	año de ocurrencia	Año ocurrencia	Año de ocurrencia	año de ocurrencia	año de ocurrencia	Año ocurrencia	Año de ocurrencia	Año ocurrencia	año_ocu	Posibles variables inútiles
dia de ocu	dia de ocurrencia	dia de ocu	Dia de ocurrencia	dia de ocurrencia	dia de ocurrencia	dia de ocu	dia de ocurrencia	dia de ocu	dia_ocu	
Dia de la semaa	dia de la semana	Dia de la semana	dia semana	Dia de la semana	dia de semana	Dia de la semaa	dia de la semana	Dia de la semaa	dia_sem_ocu	
Mes de ocurrencia	Mes de ocurrencia	Mes	Mes	mes	mes	Mes de ocurrencia	mes	Mes de ocurrencia	mes_ocu	
hora de ocurrencia	hora de ocurrencia	hora de ocurrencia	hora de ocurrencia	hora	hora	hora de ocurrencia	hora	hora de ocurrencia	hora_ocu	
g hora	g hora	g hora	g hora	g hora	g hora	g hora	g hora	g hora	g_hora	
g hora 5	g hora 5	g hora 5	g hora 5	g hora 5	g hora 5	g hora 5	g hora 5	g hora 5	g_hora_5	
Municipio	Municipio	Municipio	Municipio	Municipio	Municipio	Municipio	Municipio	Municipio	mun_ocu	
Departamento	Departamento	Departamento	Departamento	Departamento	Departamento	Departamento	Departamento	Departamento	dep_ocu	
Area geografica	area geografica	area geografica	Area geografica	area geografica	area geografica	Area geografica	area geografica	Area geografica	area_geo	
Zona	zona	zona	Zona	zona de ocurrencia	zona	Zona	zona	Zona	zona_ocu	
sexo de la persona	Sexo	sexo	sexo conductor	sexo	sexo piloto	sexo de la persona	sexo piloto	sexo de la persona	sexo_pil	
edad	Edad	edad	edad del conductor	edad	edad piloto	edad	edad	edad	edad_pil	
edad-80	edad-80	edad 80	edad-80	edad-80	edad-80	edad-80	edad-80	edad-80	edad_80	
edad-60	edad-60	edad 60	edad-60	edad-60	edad-60	edad-60	edad-60	edad-60	edad_60	
edad quincenal	edad quincenal	edad 15	edad quincenal	edad quincenal	edad quincenal	edad quincenal	edad quincenal	edad quincenal	edad_15	
mayor	mayor	Mayor	mayor	mayor o menor	mayor	mayor	mayor	mayor	mayor	
grupo edad piloto	grupo edad piloto	grupo edad piloto	grupo edad piloto	grupo edad piloto	grupo edad piloto	grupo edad piloto	grupo edad piloto	grupo edad piloto	g_edad	
edad m1	edad m1	edad m1	edad m1	edad m1	edad m1	edad m1	edad m1	edad m1	edad_m1	
estado del piloto	estado del piloto	estado del conductor	estado del conductor	estado piloto	estado del piloto	estado del piloto	estado piloto	estado del piloto	estado_pil	
tipo de vehículo	tipo de vehículo	tipo	tipo	tipo de vehículo	tipo vehículo	tipo	tipo	tipo de vehículo	tipo_veh	
Marca del vehículo	marca	marca	marca	marca	marca	Marca del vehículo	marca	Marca del vehículo	marca_veh	
color del vehículo	color	color	color	color	color	color del vehículo	color	color del vehículo	color_veh	
modelo	modelo	modelo	modelo	modelo	modelo	modelo	modelo	modelo	modelo_veh	
grupo modelo	grupo modelo	grupo modelo	Grupo modelo	grupo modelo	grupo modelo	grupo modelo	grupo modelo	grupo modelo	g_modelo_veh	
causa del accidente	causa del accidente	causa del accidente	causa del accidente	causa del accidente	causa del accidente	causa del accidente	causa del accidente	causa del accidente	causa_acc	
Tipo de evento	tipo de evento	tipo de evento	tipo de evento	tipo de evento	tipo de evento	Tipo de evento	tipo de evento	Tipo de evento	tipo_eve	

- Fallecidos y lesionados

2017	2016	2015	2014	2013	2012	2011	2010	2009		
mes de ocurrencia	mes de ocurrencia	mes de ocurrencia	Mes de ocurrencia	Mes de ocurrencia	Mes de ocurrencia	Mes de ocurrencia	Mes de ocurrencia	Mes de ocurrencia		agregado
día de ocurrencia	Día de la semana de ocurrencia	día de ocurrencia	día de la semana	Día de la semana	Día de la semana	Día de la semana	Día de la semana	Día de la semana		cambio de nombre
grupo de hora	Grupo de hora	grupo de hora	Grupo de hora	Grupo de edad	Grupo de edad	Grupo de edad	Grupo de edad	Grupo de edad		están en todos
grupo de jornada	Grupo de hora mañana, tarde y noche	grupo de jornada	grupo de jornada	grupo de jornada	grupo de jornada	grupo de jornada	grupo de jornada	grupo de jornada		no están en todos
Departamento	Departamento	departamento	Departamento	Departamento	Departamento	Departamento	Departamento	Departamento		
municipio	Municipio de ocurrencia	municipio	Municipio de ocurrencia	Municipio de ocurrencia	Municipio de ocurrencia	Municipio de ocurrencia	Municipio de ocurrencia	Municipio de ocurrencia		
zona	zona de ocurrencia	zona de ocurrencia	Zona de ocurrencia	Zona de ocurrencia	Zona de ocurrencia	Zona de ocurrencia	Zona de ocurrencia	Zona de ocurrencia		
area geografica	área geográfica	área geográfica	Área geográfica	Área geográfica	Área geográfica	Área geográfica	Área geográfica	Área geográfica		
sexo	sexo de la persona	sexo	Sexo de la víctima	Sexo	Sexo	Sexo	Sexo	Sexo		
edad	edad de la persona	edad	Edad de la víctima	Edad 1	Edad 1	Edad 1	Edad 1	Edad 1		
mayor o menor de edad	mayor o menor	mayor o menor	Mayor o menor	mayor o menos	mayor o menos	mayor o menos	mayor o menos	mayor o menos		
grupo de edad 60	grupo de edad 60 y más	grupo de edad	grupo de edad	Grupo de edad	Grupo de edad	Grupo de edad	Grupo de edad	Grupo de edad		
grupo de edad 80	grupo de edad 80 y más	grupo edad 80 o mas	grupo edad 80 o mas	grupo edad 80 o mas	grupo edad 80 o mas	grupo edad 80 o mas	grupo edad 80 o mas	grupo edad 80 o mas		
edades quinquenales	Edades quinquenales	edades quinquenales	edades quinquenales	edades quinquenales	edades quinquenales	edades quinquenales	edades quinquenales	edades quinquenales		
fallecido o lesionado	Fallecidos y lesionados	fallecido o lesionado	Fallecido o Lesionado	Fallecido o Lesionado	Fallecido o Lesionado	Fallecido o Lesionado	Fallecido o Lesionado	Fallecido o Lesionado		
intemado o no intemado	Intemado o no intemado	intemado o no intemado	intemado o no intemado	intemado o no intemado	intemado o no intemado	intemado o no intemado	intemado o no intemado	intemado o no intemado		
tipo de vehículo	estado del conductor	estado del conductor	Estado del conductor	Estado del implicado	Estado del implicado	Estado del implicado	Estado del implicado	Estado del implicado		
marca del vehículo	tipo de vehículo	tipo de vehículo	Tipo de vehículo	Tipo de vehículo	Tipo de vehículo	Tipo de vehículo	Tipo de vehículo	Tipo de vehículo		
color de vehículo	Marca de vehículo	marca del vehículo	Marca del vehículo	marca del vehículo	marca del vehículo	marca del vehículo	marca del vehículo	marca del vehículo		
modelo del vehículo	Color del vehículo	color de vehículo	Color del vehículo	Color del vehículo	Color del vehículo	Color del vehículo	Color del vehículo	Color del vehículo		
grupo modelo del vehículo	Modelo del vehículo	modelo del vehículo	Modelo del vehículo	modelo del vehículo	modelo del vehículo	modelo del vehículo	modelo del vehículo	modelo del vehículo		
tipo de evento	tipo de evento	tipo de evento	Tipo de evento	tipo de evento	tipo de evento	tipo de evento	tipo de evento	tipo de evento		

❖ Columnas agregadas

- En los archivos anteriormente mostrados se indica cuales son las columnas que originalmente no están en ellos, por lo que en este paso se agrega a cada año y archivo correspondiente a fallecidos y lesionados, vehículos involucrados y hechos de tránsito los que les hacen falta para poder unificarlos en una sola base de datos.
- Algunos de los archivos carecían del año en el que fueron ocasionados los accidentes por lo que se le agregaron variables con años de donde fue sacada dicha información.
- Al terminar de agregar las columnas se corrobora que todos los documentos según al tema que correspondan tengan la misma cantidad de variables.

❖ Renombramiento de variables

- Para poder usar una función llamada rbind se necesita tener las misma cantidad de columnas y con el mismo nombre para hacer una unión de ellas. Por lo que se hizo para cada parte que conforman los accidentes de tránsito.

❖ Unión de las tres partes

- Al unir los 3 archivos con la información de todos los años, le agregamos una variables extra para saber de qué tipo de archivo era. Con esto logramos identificar de donde originalmente pertenecen los archivos.

El procedimiento detallado del código se encuentra en la documentación del mismo.