

Proyecto Siniestros en CABA

Equipo:

Ariana Diaz

Contenido

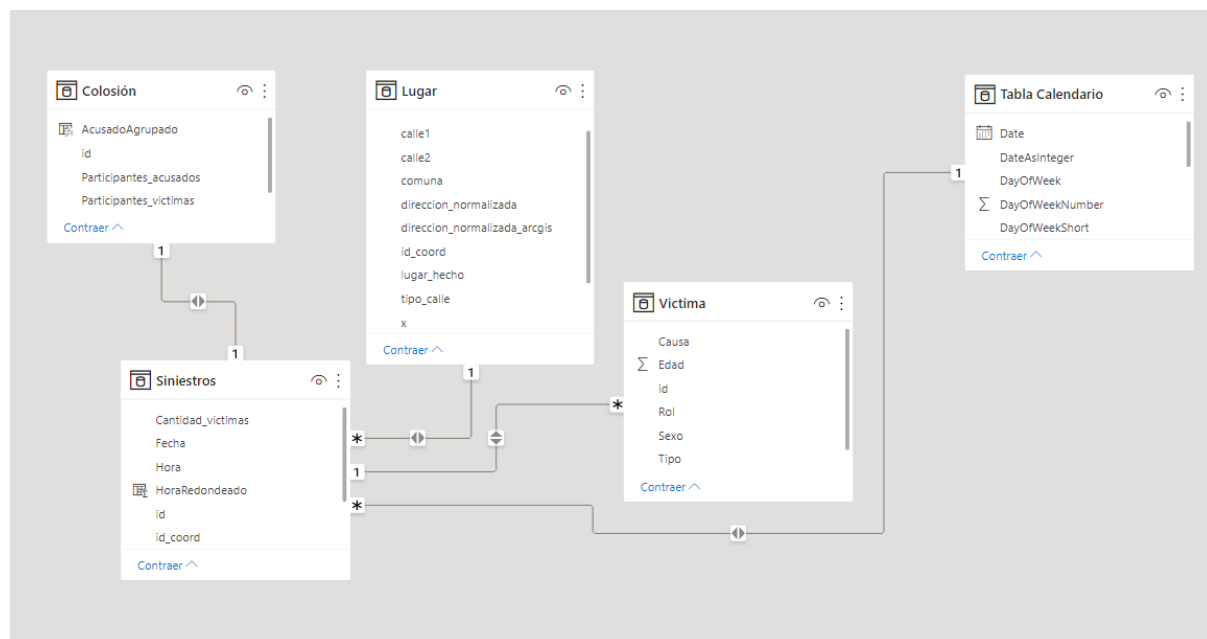
Base de Datos	3
1. Breve Descripción de la Temática de los datos	3
2. Diagrama Entidad-Relación de la Base de Datos	3
3. Descripción de las Tablas	3
Visualización	5
1. Objetivo	5
2. Transformación de datos	5
3. Análisis Funcional del Tablero	5
Medidas Calculadas	7
Conclusiones	8

Base de Datos

1. Breve Descripción de la Temática de los datos

Se cuenta con un dataset en formato .csv del gobierno de la Ciudad de Buenos Aires con los siniestros ocurridos entre 2015 y 2018. Contiene datos de la ubicación de los accidentes, víctimas, entre que vehículos fue, etc.

2. Diagrama Entidad-Relación de la Base de Datos



3. Descripción de las Tablas

Para facilitar el tratamiento de datos se dividió la información en distintas tablas y se agregó un id para los siniestros:

Tabla Siniestros:

Tipo de clave	Campo	Tipo de campo
PK-index	id	int
-	Fecha	date
-	Hora	int
-	Mes	int
-	Fecha	date
-	Cantidad_victimas	int
FK	Id_coord	int

Tabla Lugar:

Tipo de clave	Campo	Tipo de campo
PK	Id_coord	int
-	X	int
-	Y	int
-	Comuna	Int
-	Lugar_hecho	text
-	Dirección_normalizada	text
-	Tipo_calle	text
-	direccion_normalizada_a_rgis	text
-	Calle1	text
-	Altura	text
-	Calle2	text
-	X_corregido	int
-	Y_corregido	int

Se generó un id con la unión entre los primeros 6 dígitos de las coordenadas. De esta forma se pudieron eliminar los duplicados y que cada id_coord tuviera una dirección normalizada, a diferencia de antes, donde cada id podía tener una dirección distinta (aunque igual).

Por ejemplo, bajo el mismo id_coord la dirección puede ser Belgrano & Huergo o Huergo & Belgrano.

No se utilizaron las siguientes columnas: calle1, calle2, altura, lugar_hecho, comuna y dirección_normalizada.

Tabla Víctima:

Tipo de clave	Campo	Tipo de campo
FK	id	text
-	Causa	text
-	Rol	text
-	Tipo	text
-	Sexo	texto
-	Edad	int

Cabe destacar que no siempre se cuentan con los datos de todas las víctimas de un siniestro.

No se utilizó la columna edad.

Tabla Colisión:

Tipo de clave	Campo	Tipo de campo
FK	id	int
-	Tipo_colision	text
-	Participantes_victimas	text
-	Participantes_acusados	text

Visualización

1. Objetivo

Se analizarán los siniestros ocurridos en CABA, su distribución por fecha, tipos de accidentes, de víctimas y acusados. Se identificarán los días, hora y lugares con mayores siniestros. Para esto último se utiliza un mapa para facilitar su visualización.

2. Transformación de datos

En este apartado se detallarán los cambios realizados en los datos importados:

Siniestros

- La columna Fecha cuenta con formato MM/DD/YY. Para poder utilizar estos datos correctamente se cambió la *Configuración Regional a inglés (Estados Unidos)*.
- La columna Hora se importó con el siguiente formato: 31/12/1899 hh:mm:ss. Se extrajeron los últimos 8 caracteres para contar con la información de la hora únicamente y se cambió a tipo de dato Hora.
- Se creó la columna RangoHorario para terminar si el horario en que ocurrió el siniestro fue durante la mañana, tarde o noche. Este dato no fue utilizado en ninguna solapa, dado que la distribución era similar para los tres momentos del día.

Lugar

- Se dividieron las columnas x_corregido e y_corregido por 1.000 para obtener el formato de coordenadas. Luego se cambió el tipo de dato a latitud y longitud respectivamente.
- Pese a tener la misma coordenada, muchos lugares contenían distinto nombre de calle e intersección, en el Excel se modificó esto para poder las coordenadas tengan los datos correctos de la dirección.

Victima

- Se creó una columna TipoAgrupado, con el tipo de víctima agrupado en los principales 4 grupos para facilitar su gráfico.

Tabla Calendario

- Se creó esta tabla con los datos de la fecha de los siniestros.

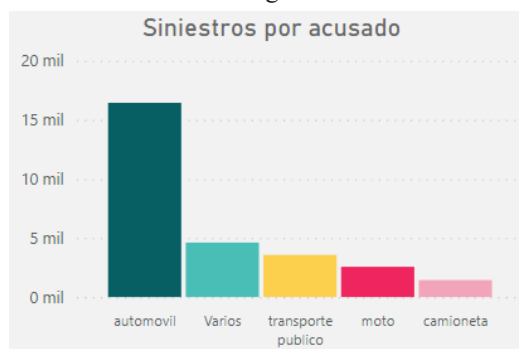
3. Análisis Funcional del Tablero

En este apartado se explicará la información incluida en cada solapa.

Página 1:

Contiene:

- Tarjeta 1, #Siniestros: Cantidad de Siniestros contando los id de la Tabla Siniestros
- Tarjeta 2, #Víctimas: Cantidad de Víctimas utilizando la sumatoria de la columna Cantidad_Victimas de la Tabla Siniestros.
- Gráfico de columnas 1, Siniestros por tipo de accidente: Cuenta con un top 5 de los tipos de colisiones en los siniestros.
- Gráfico de líneas, Siniestros vs Tiempo: Distribución de los siniestros en el tiempo. Utilizando los siguientes botones del gráfico se permite ver los accidentes por meses en un determinado año, por meses en general, etc.
- Gráfico de torta, Víctimas por género: Se utilizó la columna Sexo de la Tabla Victimas para identificar el género de las víctimas. Menos de un 1% no contenía este dato por lo que no figura en el gráfico.
- Gráfico de columnas 2, Siniestros por rol de víctima: Cuenta con un top 5 con los principales roles que tienen las víctimas por siniestros. Se accede a este gráfico con el botón “Víctima”
- Gráfico de columnas 3, Siniestros por acusado: Cuenta con un top 5 con los principales acusados en los siniestros. Se accede a este gráfico con el botón “Acusado”.



- También se cuentan con los siguientes filtros: Año, Mes, Causa. Estos afectan a todos los elementos de la página excepto al gráfico de líneas.

Página 2:



Contiene

- Mapa: Cuenta con los 10 lugares donde ocurrieron más siniestros.
- Etiqueta 1, Día donde ocurren más siniestros: Se utilizó la medida ModaDia.
- Etiqueta 2, Hora donde ocurren más siniestros: Se utilizó la medida ModaHora
- Tabla: Cuenta con los 10 lugares donde ocurrieron más siniestros y la cantidad de siniestros por año (utilizando la medida SiniestrosXAño). Si se hace clic en las direcciones se puede ver su ubicación en el mapa.
- Gráfico de barras apiladas, Tipo de calle por causa: Distribución del tipo de causa por el tipo de calle. Se empleó la columna tipo_calle de la Tabla lugar, y la columna Causa de la tabla Víctima.
- Gráfico de torta, Víctima fatal por rol: Se emplearon los datos de la columna TipoAgrupado de la Tabla Víctima. También filtro por Causa “homicidio”.

Medidas Calculadas

- ModaHora: se calculó la hora en la cual ocurren más siniestros. Se utilizó la columna auxiliar HoraRedondeado para utilizar en esta fórmula.
- ModaDia: se calculó el día de la semana en el cual ocurren más siniestros.
- SiniestrosxAño: Promedio de Siniestros por año, considerando la cantidad total de siniestros en el período 2015-2018 (4 años).
- CantidadSiniestros: Con la función Count se contabilizó los id para conocer la cantidad total.

Conclusiones

- Gran disminución de siniestros en 2018.
- Las motos son las que tienen mayor participación en los siniestros, mayoritariamente como víctimas. Tiene sentido debido a su vulnerabilidad de la persona en comparación con otros transportes, y a la falta de carriles para este tipo de transporte y probable imprudencia de los conductores de estos vehículos.
- Los automóviles son los mayores acusados como generadores del siniestro, son la mayoría de vehículos y tienen gran poder de daño.
- Los días donde ocurren más siniestros los domingos, concuerdan con los días donde mucha gente decide aprovechar el fin de semana y recorre mayores distancias dentro de la ciudad.
- La hora donde ocurren más siniestros es a las 18:00hs, hora pico durante los días de semana y donde se da el mayor flujo de vehículos.
- Pese a que la menor cantidad de siniestros se dan en autopistas, tienen mayor fatalidad que en avenida o calle.
- La mayor cantidad de víctimas fatales son pasajeros/ conductores de moto o peatones, debido a que son los más vulnerables físicamente ante un accidente.