

ÍNDICE DECISIONES POWER BI

ETAPA PREPARACIÓN DE DATOS:	2
CHARACTERISTICS	2
PLACES	3
USERS	4
VEHICLES	5
ETAPA MODELADO: TABLAS AUXILIARES	5
Día Semana (Columna de Characteristics y Tabla)	5
Sexo (Columna de Users y Tabla)	5
Grupo Etario (Columna de Users y Tabla)	5
GPS:	6
Gravedad:	6
Uso Eq Seguridad:	6
Motivo Viaje:	6
Categoría Camino:	7
Categoría Usuario:	7
Categoría Vehículo:	7
Condición Atmosférica:	7
Condición Superficie:	7
Mes:	7
ETAPA MODELADO: MEDIDAS	8
Fecha Min y Max:	8
D COUNT NUM ACC:	8
Num Acc W Blank:	8
COUNT NUM ACC:	8
Num Acc U W Blank:	8

DECISIONES POWER BI

ETAPA PREPARACIÓN DE DATOS:

GENERAL:

1. Los errores de tipo, usualmente por el uso de NA, son reemplazados por *null*.
2. Considero que el reporte está destinado a un público de habla hispana no especializada en tecnicismos propios del tema. Por lo que se utilizan headers de columnas en español y más descriptivas que por default.
3. Por la gran cantidad de filas en las tablas principales, se prioriza que estén normalizadas. Por lo que voy a tener un número o caracter en una tabla principal, que va a hacer referencia (si el reporte lo amerita), a una tabla secundaria con los mismos Id y la descripción de cada uno; evitando así mayor consumo de recursos.

CARACTERISTICS

- **long:**
 - replace errors: Si error, reemplazar "NA" por "null" debido a formato incorrecto.
 - fix: Le agrego un punto luego del primer caracter. Para esto los separo en dos casos, aquellos que tienen '-' y los que no. Los trato de manera distinta y luego los uno.
 - change type: Alfanumérico -> Decimal
- **lat:**
 - replace errors: Si error, reemplazar "NA" por "null" debido a formato incorrecto.
 - fix: Le agrego un punto luego de los dos primeros caracteres.
 - change type: Alfanumérico -> Decimal
 - Filter: En la visualización del mapa filtro los valores correspondientes a Europa. El rango definido es entre 40 y 52 ($40 \leq lat \leq 52$)
- **Clima(atm):**
 - replace errors: Si error, reemplazar "NA" por "null" debido a formato incorrecto.
 - rename: atm -> Clima (condición atmosférica)
 - replace: null -> 9 (Otros)
- **Tipo Colisión(col):**
 - replace errors: Si error, reemplazar "NA" por "null" debido a formato incorrecto.
 - rename: col -> Tipo_Colisión
- **Comuna (com):**
 - replace errors: Si error, reemplazar "NA" por "null" debido a formato incorrecto.
 - rename: com -> Comuna

- **Año (an):**
 - rename: an -> Año
 - sum: Le añadido 2000 para que quede con el año correcto
- **Mes (mois):** rename: mois -> Mes
- **Día (jour):** rename: jour -> Día
- **Hora y Minuto (hrmn):**
 - split: Los separo en 2 columnas. Tomo 2 caracteres de derecha a izquierda
 - i. Para los casos en que al separar la columna de “hora” queda en blanco por ser 00, se reemplaza el *blank* por 00.
 - rename: hr -> Hora
 - rename: mn -> Minuto
- **Iluminación (lum):**
 - rename: lum -> Iluminación
- **Área (agg):**
 - rename: agg -> Área
- **Tipo_Intersección (int):**
 - rename: int -> Tipo_Intersección
- **Dirección (adr):**
 - rename: adr -> Dirección
 - replace errors: Si error, reemplazar “NA” por “null” debido a formato
- **Código_GPS (GPS):**
 - rename: GPS -> Código_GPS
 - replaced values: [0, P, C, S, T, null] -> O
- **Departamento (dep):**
 - rename: dep -> Departamento
- **Día_Semana:**
 - Nueva columna para resolver el punto G.
 - Proceso explicado en ETAPA DE MODELADO.
- **Fecha:**
 - Merge de Dia/Mes/Año
 - Para proveer filtro a las secciones.

PLACES

- **Categoría_Camino (catr)**
 - rename
- **Número_Camino (voie)**
 - rename
 - replaced: NA->null
- **Id_Numérico_Camino (V1)**
 - rename
- **Id_Alfanum_Camino (V2)**
 - rename
- **Régimen_Tráfico (circ)**
 - rename
- **Núm_Carriles (nbv)**
 - rename
- **Carril_Reservado(vosp)**

- rename
- **Pendiente(Prof)**
 - rename
- **Plano(plan)**
 - rename
- **Ancho_Pavimento (lartpc)**
 - rename
- **Ancho_Calzada (larrout)**
 - rename
 - replaced: -81 -> null
- **Condición_Superficie (surf)**
 - rename
 - replaced: [0, null] -> 9 (Otros)
- **Infraestructura (infra)**
 - rename
- **Situación (situ)**
 - rename
- **Prox_Escuela (env1)**
 - rename

USERS

- **Ubicación_Vehículo (place):**
 - rename
 - replaced: NA -> null
- **Categoría (catu):**
 - rename
- **Gravedad (grav):**
 - rename
- **Sexo (sex):**
 - rename
- **Motivo_Viaje (trajet):**
 - rename
 - replaced: [0,null] -> 9 (de valor indefinido (0, null) a Otros (9))
- **Equipo_Seguridad y Uso_Equipo_Seguridad (secu):**
 - replaced: NA->null
 - split: 1 caracter:
 - 1er caracter: Equipo_Seguridad
 - 2do caracter: Uso_Equipo_Seguridad
 - replaced: [0, null] -> 3 (los 0 y null pasan a ser indeterminados)
 - rename
- **Ubicación_Peatón (locp):**
 - rename
- **Acción_Peatón (actp):**
 - rename
- **Compañía_Peaton (etatp):**

- rename
- **Año_Nacimiento (an_nais):**
 - rename

VEHICLES

- **Dirección_Flujo (senc)**
 - rename
- **Categoría_Vehículo (catv)**
 - rename
 - replaced: [15,16,17,18,19,20,21,30,31,32,33,34,35,36,37,38,39,40,99] -> 14 ("Otros") para proveer una descripción en común con el código 14 al momento de modelar.

ETAPA MODELADO: TABLAS AUXILIARES

Día_Semana (Columna de Characteristics y Tabla)

- Extraído en 'characteristics' luego de:
 - Hacer un merge (dd/mm/aaaa).
 - Cambiarle formato a fecha local (es.ar).
 - Extraer el número de día de la semana, donde 0 es domingo y 6 es sábado
 - Sumarle 1.
 - Asociar a nueva tabla Día_Semana, que contiene un campo Id que va a matchear con cada número extraído de la fecha. Cada id contiene en su descripción el nombre del día correspondiente, donde 1 es domingo y 7 es sábado.
- Modelado para resolver el punto G.
- Se creó una relación muchos a uno desde la Tabla[characteristics: Día_Semana] a la Tabla[Día_Semana: Id].

Sexo (Columna de Users y Tabla)

- Modelado para resolver el punto A. y F. con el fin de disponer de una descripción en vez de un mero número.
- Se creó una relación muchos a uno desde la Tabla[users: Sexo] a la Tabla[Sexo: Id].

Grupo_Etario (Columna de Users y Tabla)

- En users es una columna extraída luego de:
 - Generar la columna Edad luego de restar:
 - RELATED(characteristics(Año)) a
 - users(Año_Nacimiento).
 - Sobre la columna Edad obtenida se realizaron varios ifs concatenados donde:

- Si Edad <= 15, se le asigna 1,
 - Si Edad <= 30, se le asigna 2,
 - Si Edad <=45, se le asigna 3,
 - Si Edad <= 60, se le asigna 4,
 - Resto, se le asigna 5.
- Se creo una relación muchos a uno desde Tabla[users] a Tabla[Grupo_Etario].
- Se modeló la tabla con los lds y descripciones correspondientes mencionados en el lf concatenado.

GPS:

- Modelado para proveer descripción a cada código de GPS del punto B.
- Para los códigos no descriptos en el diccionario de datos, sus valores fueron reemplazados por "O" con la correspondiente descripción "Otros". Entonces, la tabla quedo de la siguiente manera:
 - M: Métropole.
 - A: Antilles.
 - G: Guyane.
 - R: Réunion.
 - Y: Mayotte.
 - O: Otros (null, 0, P, C, S, T).
- Se creó una relación muchos a uno desde la Tabla[caracteristics: Código_GPS] a la Tabla[GPS: Id].

Gravedad:

- Modelado para proveer descripción a los códigos de gravedad de los puntos D. y E.
- Se creó una relación muchos a uno desde la Tabla[users: Gravedad] a la Tabla[Gravedad: Id].

Uso_Eq_Seguridad:

- Modelado para proveer descripción a los códigos de uso de equipo de seguridad del punto E. donde:
 - 1: Si
 - 2: No
 - 3: Indeterminado (se reemplazaron los 0 y null por 3)
- Se creó una relación muchos a uno desde la Tabla[users: Uso_Equipo_Seguridad] a la Tabla[Uso_Eq_Seguridad: Id].

Motivo_Viaje:

- Modelado para proveer descripción a los códigos de motivo de viaje. Me parecio interesante utilizar los datos para modelar una visualización fuera de los puntos del desafio en la sección de USUARIO.
- Se creó una relación muchos a uno desde la Tabla[users: Motivo_Viaje] a la Tabla[Motivo_Viaje: Id].

Categoría_Camino:

- Modelado para proveer de descripción a los códigos de la visualización del punto C.
- Se creó una relación muchos a uno desde la Tabla[places: Categoría_Camino] a la Tabla[Categoría_Camino: Id].

Categoría_Usuario:

- Modelado para ser utilizado como filtro de las visualizaciones y proveer descripción en vez de un código.
- Se creó una relación muchos a uno desde la Tabla[users: Categoría] a la Tabla[Categoría_Usuario: Id].

Categoría_Vehículo:

- Modelado para ser utilizado como filtro de las visualizaciones y proveer descripción en vez de un código.
- Se creó una relación muchos a uno desde la Tabla[vehicles: Categoría_Vehículo] a la Tabla[Categoría_Vehículo: Id].

Condición_Atmosférica:

- Modelado para ser utilizado como filtro de las visualizaciones y proveer descripción en vez de un código.
- Se creó una relación muchos a uno desde la Tabla[characteristics: Clima] a la Tabla[Condición_Atmosférica: Id].

Condición_Superficie:

- Modelado para ser utilizado como filtro de las visualizaciones y proveer descripción en vez de un código.
- Se creó una relación muchos a uno desde la Tabla[places: Condición_Superficie] a la Tabla[Condición_Superficie: Id].

Mes:

- Proveer de descripción a cada mes en la visualización correspondiente al punto B.
- Valores 1-12 como Id con su correspondiente Descripción, donde:
 - 1: Enero
 - 2: Febrero
 -
 - 12: Diciembre
- Se creó una relación muchos a uno desde Tabla[characteristics: Mes] a Tabla[Mes: Id].

ETAPA MODELADO: MEDIDAS

Fecha Min y Max:

- Toma todas las fechas de characteristics y devuelve un texto en el siguiente formato:
MIN("dd/mm/aaaa") - MAX("dd/mm/aaaa").
 - EJ: 01/01/2005 - 31/12/2016
- Utilizado para proveer información adicional en HomePage.

D_COUNT_NUM_ACC:

- Devuelve la cantidad de accidentes distintos en characteristics.

Num_Acc_W_Blank:

- En characteristics.
- Utilizado junto con "D_COUNT_NUM_ACC", para el caso en que luego de aplicar ciertos filtros la cantidad de accidentes sea 0. Muestra 0 en vez de Blank.

COUNT_NUM_ACC:

- Devuelve la cantidad de usuarios accidentados en users.

Num_Acc_U_W_Blank:

- En users.
- Utilizado junto con "COUNT_NUM_ACC", para el caso en que luego de aplicar ciertos filtros la cantidad de usuarios accidentados sea 0. Muestra 0 en vez de Blank.