



Reto No. 3

Seguridad en las vías

OBJETIVOS

Se desea diseñar un Tipo Abstracto de Datos (TAD) para modelar un analizador de tránsito.

Esta herramienta le servirá a quienes deben decidir dónde ubicar señales de tránsito para reducir accidentes, analizar zonas geográficas más propensas a los accidentes e identificar épocas del año en las que se producen más accidentes, entre otros reportes.

Ustedes deben utilizar las estructuras de datos vistas hasta el momento, particularmente árboles binarios, para responder las consultas realizadas en el menor tiempo posible.

DINÁMICA DE TRABAJO

Se trabajará de manera grupal e individual en el reto.

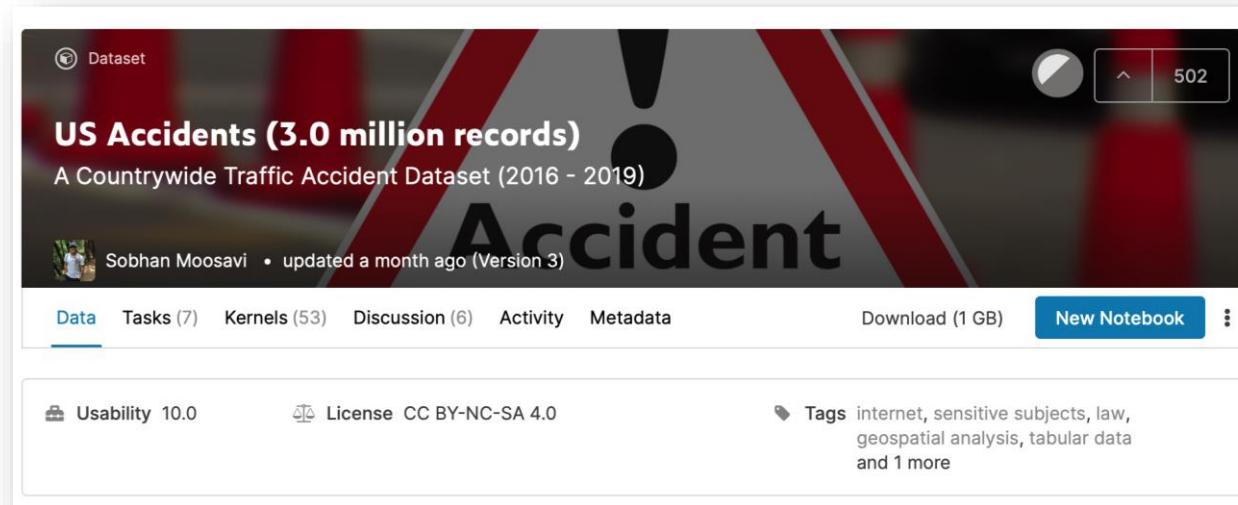
- El requerimiento No.1 se desarrollará en el laboratorio.
- Cada estudiante desarrollará **individualmente** uno de los requerimientos, seleccionándolo de entre los No. **2, 3, y 4**.
- El requerimiento **No. 5** se implementará en **grupo**.
- Los requerimientos No 6 y 7 son opcionales y si los implementan cuentan como bono.

FUENTE DE DATOS

El conjunto de datos original proviene del Proyecto US Accidents (Version 3) (URL: <https://www.kaggle.com/sobhanmoosavi/us-accidents/version/3>)

Creados y utilizados originalmente en los artículos:

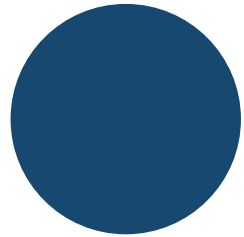
- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. "[A Countrywide Traffic Accident Dataset](#).", 2019.
- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. "[Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights](#)." In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.



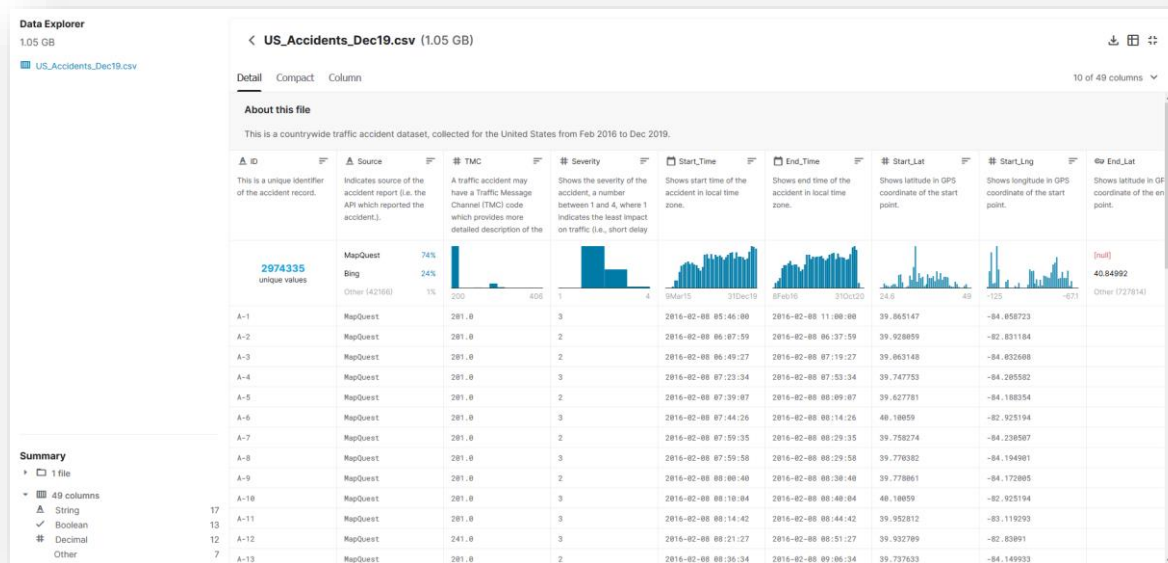
Tomado de US Accidents (Version 3), URL: <https://www.kaggle.com/sobhanmoosavi/us-accidents/version/3>

En este reto vamos a trabajar con un conjunto de datos de Kaggle, que contiene información de accidentes de tránsito en Estados Unidos entre los años 2016 y 2019.

En esta información se encuentra la fecha en la que ocurrió el accidente, la localización geográfica (latitud / longitud), la distancia afectada por el accidente y la duración del mismo, entre otros datos.



Detalles de la fuente de datos



Tomado de US Accidents (Version 3), URL:

<https://www.kaggle.com/sobhanmoosavi/us-accidents/version/3>

Los datos constan de **49** columnas:

- **17** de ellos son cadenas de caracteres o **String** (ej.: ID y Source).
- **13** son **Booleanos** (ej.: Station y Stop).
- **12** son de tipo **numerico** (ej.: TMC y Distance(ml)).
- **7** de **otros** tipos (ej.: Weather_Timestamp y Start_Time).

Algunos datos en las columnas pueden ser nulos.

Detalles de la fuente de datos

#	Attribute	Description	Type	Nullable
1	ID	This is a unique identifier of the accident record.	String	No
2	Source	Indicates source of the accident report (i.e. the API which reported the accident.).	String	No
3	TMC	A traffic accident may have a Traffic Message Channel (TMC) code which provides more detailed description of the event.	Decimal	Yes
4	Severity	Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay).	Decimal	No
5	Start_Time	Shows start time of the accident in local time zone.	Other	No
6	End_Time	Shows end time of the accident in local time zone. End time here refers to when the impact of accident on traffic flow was dismissed.	Other	No
7	Start_Lat	Shows latitude in GPS coordinate of the start point.	Decimal	No
8	Start_Lng	Shows longitude in GPS coordinate of the start point.	Decimal	No
9	End_Lat	Shows latitude in GPS coordinate of the end point.	Decimal	Yes
10	End_Lng	Shows longitude in GPS coordinate of the end point.	Decimal	Yes
11	Distance(mi)	The length of the road extent affected by the accident.	Decimal	No
12	Description	Shows natural language description of the accident.	String	No
13	Number	Shows the street number in address field.	Decimal	Yes
14	Street	Shows the street name in address field.	String	Yes
15	Side	Shows the relative side of the street (Right/Left) in address field.	String	Yes
16	City	Shows the city in address field.	String	Yes

LISTA COMPLETA EN EL DOCUMENTO EXCEL ADJUNTO A ESTA PRESENTACIÓN.

Creado con la información de US Accidents (Version 3), URL: <https://www.kaggle.com/sobhanmoosavi/us-accidents/version/3> y
US-Accidents: A Countrywide Traffic Accident Dataset, URL: https://smoosavi.org/datasets/us_accidents

FICHA DE REQUERIMIENTOS

Básico

REQ. 1: Conocer los
accidentes en una
fecha (G-LAB)

Intermedio

REQ. 2: Conocer los
accidentes anteriores a
una fecha (I)

REQ. 3: Conocer los
accidentes en un rango
de fechas (I)

REQ. 4: Conocer el
estado con mas
accidentes (I)

Avanzado

Req. 5: Conocer los
accidentes por rango
de horas (G)

REQ. 6: Conocer la zona
geográfica mas
accidentada (B)

REQ. 7: Usar el
conjunto completo de
datos (B)

REQ. 1: Conocer los accidentes en una fecha (Grupal en el Lab.)

Para implementar esta funcionalidad debe utilizar árboles binarios.

Se desea conocer el total de accidentes ocurridos para una **fecha específica**, reportando la cantidad de accidentes por severidad para dicha fecha.

REQ. 2: Conocer los accidentes anteriores a una fecha (Individual)

Se desea conocer el total de accidentes ocurridos antes de una fecha específica.

La fecha se debe ingresar siguiente el formato:

(YYYY-MM-DD)

Se debe indicar el **total** de accidentes ocurridos antes de la fecha indicada y la **fecha** en la que más accidentes se reportaron.

REQ. 3: Conocer los accidentes en un rango de fechas (Individual)

Se desea conocer el **total** de accidentes ocurridos en un **rango** de fechas

El usuario ingresa una fecha inicial y una fecha final en formato **YYYY-MM-DD**

Se debe responder con el número **total** de accidentes en ese rango de fechas, indicando la **categoría** de accidentes más reportadas en dicho rango.

REQ. 4: Conocer el estado con mas accidentes (Individual)

Se desea conocer para un rango de fechas el **estado** que más accidentes tiene reportados.

El usuario ingresa una fecha inicial y una fecha final en formato: **YYYY-MM-DD**

Se debe retornar la **fecha** con más accidentes reportados.

Req. 5: Conocer los accidentes por rango de horas (Grupal)

Se desea conocer para un **rango** de **horas** dado (hora inicial y hora final), el **total** de accidentes registrados agrupados por severidad. Igualmente se desea conocer el porcentaje que ese número representa contra el total de accidentes reportados.

Note que las horas y minutos se usan en formato 00:00-23:59

Para facilitar la consulta, los accidentes se manejarán en grupos de 30 minutos.

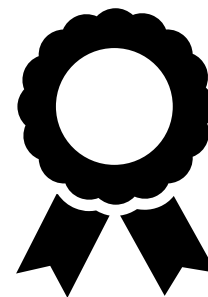
Por ejemplo:

- Número de accidentes ocurridos entre las 10:05 y 15:32 se puede aproximar a los accidentes ocurridos entre las 10:00 y las 16:00.
- Número de accidentes ocurridos entre las 14:00 y las 14:25 se puede aproximar a los accidentes ocurridos entre las 14:00 y las 14:30.

REQ. 6: Conocer la zona geográfica mas accidentada (Opcional - Bono)

Dada una latitud y una longitud, tomados como punto central, y un radio dado (por ejemplo una milla), informar cuántos accidentes en total se han producido en ese radio desde el punto de búsqueda.

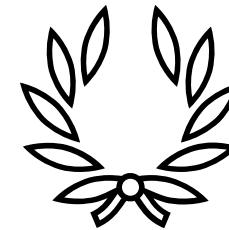
El resultado se debe presentar agrupado por el día de la semana en la que han ocurrido los accidentes y el total de accidentes reportados.



REQ. 7: Usar el conjunto completo de datos (Opcional - Bono)

- Dado el tamaño de la fuente de datos, se sugiere probar las consultas anteriores, con los datos de solo un año (2016, 2017, 2018 o 2019).
- Habrá **un segundo bono** a quienes respondan las consultas con el conjunto completo de datos (los 4 años)

- **Nota:** Cargar esta información puede tomar un tiempo considerable y un uso intensivo de la memoria de su computador.



RECOMENDACIONES

- Se deben utilizar tablas de símbolos ordenadas y todas las otras estructuras de datos del curso que considere apropiadas.
- Se debe crear un TAD con una conjunto de operaciones (API) para el analizador de tránsito.