Instituto Politécnico Nacional

Escuela Superior de Cómputo

Minería de Datos

Vianey Maravilla Pérez

5AM1

Roberto Zagal

DATA IMPORT

Practica 1: Exportación de datos y cobertura de espacio y tiempo.

En esta práctica se realizo una importación de los datos de incidentes viales que ocurrió en durante la segunda mitad del año 202.

Dentro de esto, se respondieron las preguntas planteadas en la práctica

Instrucciones:

1.- Descargue el dataset de incidentes viales, que corresponde al último semestre del 2020

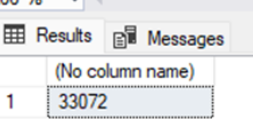
2.- Exporte el archivo CSV en el manejador de base de datos seleccionado

Desarrollo:

3.- Indique el número de registros del dataset en el manejador

Una vez hechas las instrucciones, necesitamos saber cuál es el número de registros de el dataset que vamos a utilizar, para eso debemos tener una línea de código:

***1 USE incidentes;  
2  
3 SELECT COUNT(folio) FROM dbo.incidentevial2dsem2020;***

******El código anterior nos da como resultado que el manejador de datos cuente el número de registros de incidentes viales con el folio de este mismo para así tener un total y nos ejecutará algo así:

4.- ¿Cuál es el rango de los registros del dataset en el manejador?

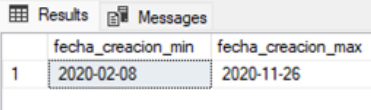
- “Fecha” (todos los relacionados)

- Latitud y Longitud

- Año\_cierre y hora\_cierre (todos los relacionados al cierre)

Fecha de creación:

***1 SELECT MIN(fecha\_creacion) AS fecha\_creacion\_min,  
2 MAX(fecha\_creacion) AS fecha\_creacion\_max  
3 FROM dbo.incidentevial2dsem2020;***

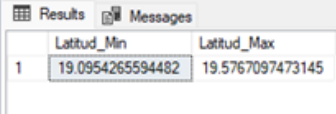
******

Fecha de Cierre

SELECT MIN(fecha\_cierre) AS fecha\_cierre\_min,  
2 MAX(fecha\_cierre) AS fecha\_cierre\_max  
3 FROM dbo.incidentevial2dsem2020;

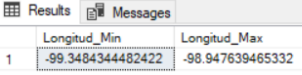


Latitud

1 SELECT MIN(latitud) AS Latitud\_Min, MAX(latitud) AS Latitud\_Max  
2 FROM dbo.incidentevial2dsem2020;

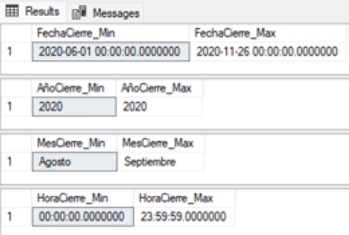
Longitud

1 SELECT MIN(longitud) AS Latitud\_Min, MAX(longitud) AS Latitud\_Max  
2 FROM dbo.incidentevial2dsem2020;



Año de cierre y hora de cierre

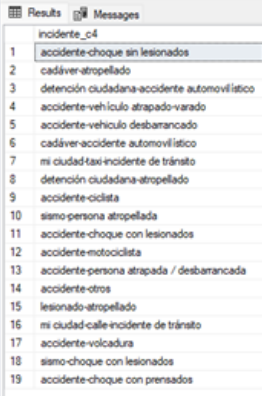
SELECT MIN(fecha\_cierre) AS FechaCierre\_Min, MAX(fecha\_cierre) AS FechaCierre\_Max  
2 FROM dbo.incidentevial2dsem2020;  
3 SELECT MIN(a~no\_cierre) AS A~noCierre\_Min, MAX(a~no\_cierre) AS A~noCierre\_Max  
4 FROM dbo.incidentevial2dsem2020;  
5 SELECT MIN(mes\_cierre) AS MesCierre\_Min, MAX(mes\_cierre) AS MesCierre\_Max  
6 FROM dbo.incidentevial2dsem2020;  
7 SELECT MIN(hora\_cierre) AS HoraCierre\_Min, MAX(hora\_cierre) AS HoraCierre\_Max  
8 FROM dbo.incidentevial2dsem2020;



5.- ¿Cuáles son los valores que toman las siguientes columnas (rango, i.e, valores posibles no repetidos) y que significado tiene (<https://datos.cdmx.gob.mx/dataset/incidentes-viales-c5>)?

- Incidente\_c4

1 SELECT DISTINCT incidente\_c4  
2 FROM dbo.incidentevial2dsem2020;



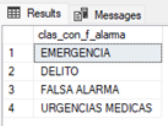
- Tipo\_entrada

1 SELECT DISTINCT tipo\_entrada  
2 FROM dbo.incidentevial2dsem2020;



- Clas\_con\_f\_alarma

SELECT DISTINCT clas\_con\_f\_alarma  
2 FROM dbo.incidentevial2dsem2020;



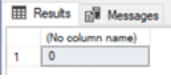
- Delegación

1 SELECT DISTINCT delegacion\_inicio  
2 FROM dbo.incidentevial2dsem2020;

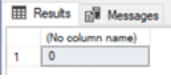
 

6.- Contar la cantidad de NULL o NULOS encontrados en las 4 columnas anteriores del punto 5

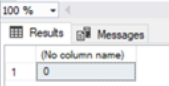
- Incidente\_c4



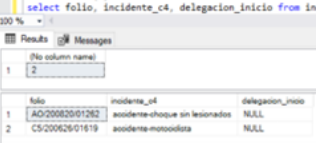
- Tipo de entrada



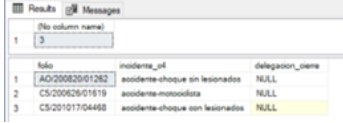
- Clase con f alarma



- Delegación de inicio



- Delegación de cierre



7.- En las columnas analizadas (del punto 4 al 6) ¿Encontró alguna anomalía en los valores? Si, sí realice una discusión de las columnas en cuestión

A la hora de la creación del análisis se lograron encontrar algunas anomalías, desde la hora, delegación de cierre, geopoint y el mes, en la hora en la creación y el cierre.

Sus periodos de los registros no corresponden al que esta establecido, pues los incidentes que se notan dentro de los registros son desde la segunda mitad del año 2020, aunque hay registros que comienzan desde febrero del mismo año, luego entonces, las fechas de cierra comenzaba desde el 1 de junio del 2020 y pues se comienza dentro de todo el día 1 de julio de 2020.

8.- Conclusiones:

Al analizar el siguiente dataset se llega a la conclusión que dentro de todos los accidentes viales y su mayor frecuencia es dentro de Iztapalapa, puesto que se deduce que es por el alta de población y vehículos puesto que esto no lo tienen delegaciones que están a sus alrededores.

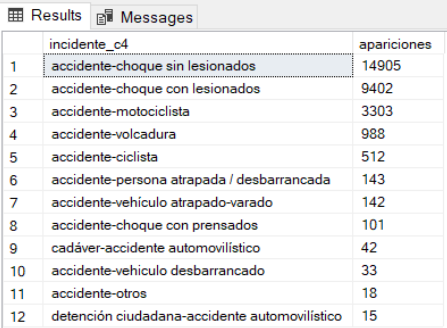
Cuando se visualizaban los registros dentro de nuestro dataset, hay momentos en donde se encuentran datos faltantes y se expresan como datos nulos, en algunas ocasiones se rellenan mediante otras columnas, pero la mayoría de tiempo no.

**CONSULTAS DENTRO DE CLASE ADICIONALES PARA AGREGARLAS A LA PRACTICA**

Accidentes:

SELECT incidente\_c4, COUNT (\*) AS apariciones

FROM dbo.incidentevial2dsem2020

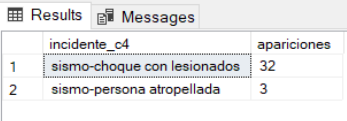
 WHERE incidente\_c4 LIKE ‘%accidente%’

GROUP BY incidente\_c4

ORDEN BY apariciones DESC;

SISMO

SELECT incidente\_c4, COUNT (\*) AS apariciones

 FROM dbo.incidentevial2dsem2020

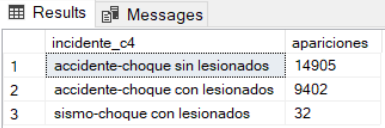
WHERE incidente\_c4 LIKE ‘%sismo%’

GROUP BY incidente\_c4

ORDEN BY apariciones DESC;

LESIONADO

SELECT incidente\_c4, COUNT (\*) AS apariciones

 FROM dbo.incidentevial2dsem2020

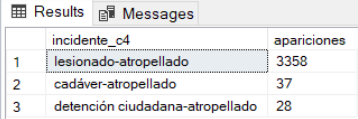
WHERE incidente\_c4 LIKE ‘%lesionado%’

GROUP BY incidente\_c4

ORDEN BY apariciones DESC;

ATROPELLADO

SELECT incidente\_c4, COUNT (\*) AS apariciones

 FROM dbo.incidentevial2dsem2020

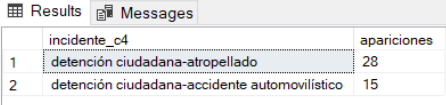
WHERE incidente\_c4 LIKE ‘%atropellado%’

GROUP BY incidente\_c4

ORDEN BY apariciones DESC;

DETENCIONES

SELECT incidente\_c4, COUNT (\*) AS apariciones

 FROM dbo.incidentevial2dsem2020

WHERE incidente\_c4 LIKE ‘%dentención%’

GROUP BY incidente\_c4

ORDEN BY apariciones DESC;