



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 realizó una importación de los datos de incidentes viales que ocurrió en durante la segunda mitad del año 2020.

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

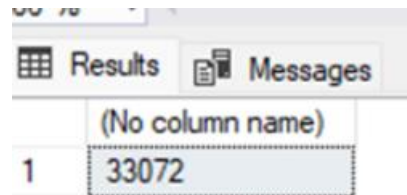
Una vez hechas las instrucciones, necesitamos saber cuál es el número de registros de el dataset que vamos a utilizar, para eso debemos tener una línea de código:

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:



	(No column name)
1	33072

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;

Results Messages		
	fecha_creacion_min	fecha_creacion_max
1	2020-02-08	2020-11-26

Fecha de Cierre

```

1 SELECT MIN(fecha_cierre) AS fecha_cierre_min,
2 MAX(fecha_cierre) AS fecha_cierre_max
3 FROM dbo.incidentevial2dsem2020;

```

Results Messages		
	fecha_cierre_min	fecha_cierre_max
1	2020-06-01 00:00:00.0000000	2020-11-26 00:00:00.0000000

Latitud

```

1 SELECT MIN(latitud) AS Latitud_Min, MAX(latitud) AS Latitud_Max
2 FROM dbo.incidentevial2dsem2020;

```

Results Messages		
	Latitud_Min	Latitud_Max
1	19.0954265594482	19.5767097473145

Longitud

```

1 SELECT MIN(longitud) AS Latitud_Min, MAX(longitud) AS Latitud_Max
2 FROM dbo.incidentevial2dsem2020;

```

Results Messages		
	Longitud_Min	Longitud_Max
1	-99.3484344482422	-98.947639465332

Año de cierre y hora de cierre

```

1 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
FROM dbo.incidentevial2dsem2020;
```

Results		Messages	
	FechaCierre_Min	FechaCierre_Max	
1	2020-06-01 00:00:00.0000000	2020-11-26 00:00:00.0000000	
	AñoCierre_Min	AñoCierre_Max	
1	2020	2020	
	MesCierre_Min	MesCierre_Max	
1	Agosto	Septiembre	
	HoraCierre_Min	HoraCierre_Max	
1	00:00:00.0000000	23:59:59.0000000	

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;
```

Results		Messages	
	incidente_c4		
1	accidente-choque sin lesionados		
2	cadáver-atropellado		
3	detención ciudadana-accidente automovilístico		
4	accidente-vehículo atrapado-varado		
5	accidente-vehículo desbarancado		
6	cadáver-accidente automovilístico		
7	mi ciudad-taxi-incidente de tránsito		
8	detención ciudadana-atropellado		
9	accidente-ciclista		
10	sismo-persona atropellada		
11	accidente-choque con lesionados		
12	accidente-motociclista		
13	accidente-persona atrapada / desbarancada		
14	accidente-otros		
15	lesionado-atropellado		
16	mi ciudad-calle-incidente de tránsito		
17	accidente-volcadura		
18	sismo-choque con lesionados		
19	accidente-choque con prensados		

- Tipo_entrada

```
1 SELECT DISTINCT tipo_entrada
2 FROM dbo.incidentevial2dsem2020;
```

	Results	Messages
	tipo_entrada	
1	BOTÓN DE AUXILIO	
2	LLAM	
3	LLAMADA APP911	
4	APLICATIVOS	
5	RADIO	
6	LLAMADA DEL 911	
7	REDES	
8	CÁMARA	

- Clas_con_f_alarma

```
SELECT DISTINCT clas_con_f_alarma
2 FROM dbo.incidentevial2dsem2020;
```

	Results	Messages
	clas_con_f_alarma	
1	EMERGENCIA	
2	DELITO	
3	FALSA ALARMA	
4	URGENCIAS MEDICAS	

- Delegación

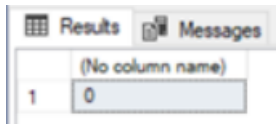
```
1 SELECT DISTINCT delegacion_inicio
2 FROM dbo.incidentevial2dsem2020;
```

	delegacion_inicio
1	MIGUEL HIDALGO
2	MILPA ALTA
3	AZCAPOTZALCO
4	CUAJIMALPA
5	TLALPAN
6	NULL
7	VENUSTIANO CARRANZA
8	MAGDALENA CONTRERAS
9	IZTACALCO
10	CUAUHTEMOC
11	XOCHIMILCO
12	TLAHUAC
13	GUSTAVO A. MADERO
14	BENITO JUAREZ
15	IZTAPALAPA
16	ALVARO OBREGON
17	COYOACAN

	delegacion_cieme
1	MIGUEL HIDALGO
2	MILPA ALTA
3	AZCAPOTZALCO
4	CUAJIMALPA
5	TLALPAN
6	NULL
7	NULL
8	VENUSTIANO CARRANZA
9	MAGDALENA CONTRERAS
10	IZTACALCO
11	CUAUHTEMOC
12	XOCHIMILCO
13	TLAHUAC
14	GUSTAVO A. MADERO
15	BENITO JUAREZ
16	IZTAPALAPA
17	ALVARO OBREGON
18	COYOACAN

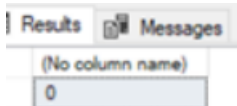
6.- Contar la cantidad de NULL o NULOS encontrados en las 4 columnas anteriores del punto 5

- Incidente_c4



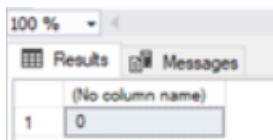
(No column name)
0

- Tipo de entrada



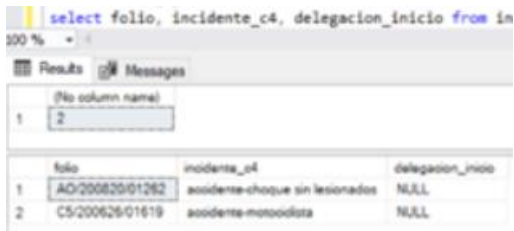
(No column name)
0

- Clase con f alarma



(No column name)
0

- Delegación de inicio



```
select folio, incidente_c4, delegacion_inicio from in
```

(No column name)
2

folio	incidente_c4	delegacion_inicio
AO/200626/01262	accidente-choque sin lesionados	NULL
CS/200626/01619	accidente-motociclista	NULL

- Delegación de cierre



(No column name)
3

folio	incidente_c4	delegacion_cierre
AO/200626/01262	accidente-choque sin lesionados	NULL
CS/200626/01619	accidente-motociclista	NULL
CS/201017/04468	accidente-choque con lesionados	NULL

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 cierre 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
```

```
ORDER BY apariciones DESC;
```

	incidente_c4	apariciones
1	accidente-choque sin lesionados	14905
2	accidente-choque con lesionados	9402
3	accidente-motociclista	3303
4	accidente-volcadura	988
5	accidente-ciclista	512
6	accidente-persona atrapada / desbarrancada	143
7	accidente-vehículo atrapado-varado	142
8	accidente-choque con prensados	101
9	cadáver-accidente automovilístico	42
10	accidente-vehículo desbarrancado	33
11	accidente-otros	18
12	detención ciudadana-accidente automovilístico	15

SISMO

```
SELECT incidente_c4, COUNT (*) AS apariciones
FROM dbo.incidentevial2dsem2020
WHERE incidente_c4 LIKE '%sismo%'
GROUP BY incidente_c4
ORDER BY apariciones DESC;
```

Results Messages		
	incidente_c4	apariciones
1	sismo-choque con lesionados	32
2	sismo-persona atropellada	3

LESIONADO

```
SELECT incidente_c4, COUNT (*) AS apariciones
FROM dbo.incidentevial2dsem2020
WHERE incidente_c4 LIKE '%lesionado%'
GROUP BY incidente_c4
ORDER BY apariciones DESC;
```

Results Messages		
	incidente_c4	apariciones
1	accidente-choque sin lesionados	14905
2	accidente-choque con lesionados	9402
3	sismo-choque con lesionados	32

ATROPELLADO

```
SELECT incidente_c4, COUNT (*) AS apariciones
FROM dbo.incidentevial2dsem2020
WHERE incidente_c4 LIKE '%atropellado%'
GROUP BY incidente_c4
ORDER BY apariciones DESC;
```

Results Messages		
	incidente_c4	apariciones
1	lesionado-atropellado	3358
2	cadáver-atropellado	37
3	detención ciudadana-atropellado	28

DETENCIONES

```
SELECT incidente_c4, COUNT (*) AS apariciones
FROM dbo.incidentevial2dsem2020
WHERE incidente_c4 LIKE '%detención%'
GROUP BY incidente_c4
ORDER BY apariciones DESC;
```

Results Messages		
	incidente_c4	apariciones
1	detención ciudadana-atropellado	28
2	detención ciudadana-accidente automovilístico	15