



Instituto Politécnico Nacional

Escuela Superior de Cómputo



Data Mining

Practica 1

Exportación de datos y cobertura de espacio y tiempo.

Grupo: 3CV18

Integrantes.

Cazares Martínez Maximiliano

Ramos Nieves Adrián

Profesor:

Roberto Eswart Zagal Flores

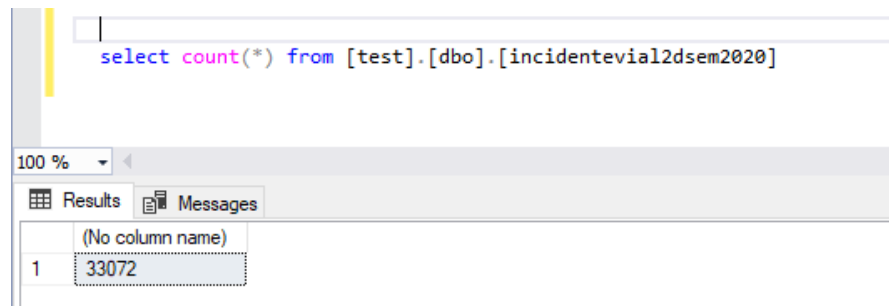
Introducción

En la presente practica analizaremos una base de datos con registros de accidentes viales en la ciudad de México en el último semestre del 2020. Teniendo 33072 registro realizamos varias consultas con el fin de analizar sin mucho detalle la información que existe en dicha base de datos.

La práctica se realizó usando Microsoft SQL Server Management Studio 2018 y Microsoft SQL Server 2019 Developer, estos dos softwares fueron descargados desde la página oficial de Microsoft y configurados mediante la opción básica que el asistente de instalación nos ofrece.

Desarrollo

El número de registros obtenidos en la base de datos fueron de 33,072, realizando la siguiente consulta en el manejador.

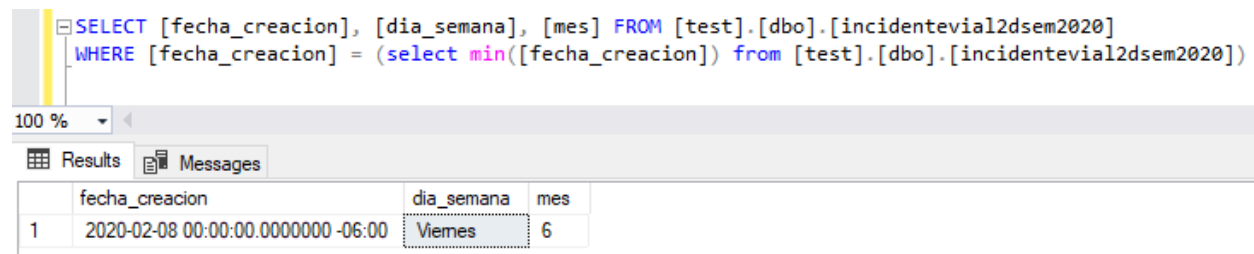


```
select count(*) from [test].[dbo].[incidentevial2dsem2020]
```

(No column name)	
1	33072

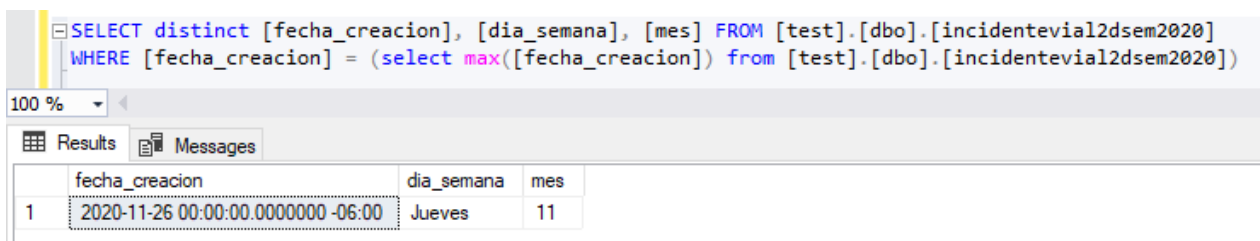
Encontramos los rangos del valor mínimo y valor máximo de los siguientes atributos:

- **Fecha:** Con la siguiente información podemos tener una visión de las fechas próximas en que ocurren los accidentes viales.



```
SELECT [fecha_creacion], [dia_semana], [mes] FROM [test].[dbo].[incidentevial2dsem2020] WHERE [fecha_creacion] = (select min([fecha_creacion]) from [test].[dbo].[incidentevial2dsem2020])
```

	fecha_creacion	dia_semana	mes
1	2020-02-08 00:00:00.0000000 -06:00	Viernes	6



```
SELECT distinct [fecha_creacion], [dia_semana], [mes] FROM [test].[dbo].[incidentevial2dsem2020] WHERE [fecha_creacion] = (select max([fecha_creacion]) from [test].[dbo].[incidentevial2dsem2020])
```

	fecha_creacion	dia_semana	mes
1	2020-11-26 00:00:00.0000000 -06:00	Jueves	11

- **Latitud.**

```
select min([latitud]) from [test].[dbo].[incidentevial2dsem2020]
```

100 %

(No column name)	
1	19.095427

```
select max([latitud]) from [test].[dbo].[incidentevial2dsem2020]
```

100 %

(No column name)	
1	19.57671

- **Longitud**

```
select min([longitud]) from [test].[dbo].[incidentevial2dsem2020]
```

100 %

(No column name)	
1	-99.348434

```
select max([longitud]) from [test].[dbo].[incidentevial2dsem2020]
```

100 %

(No column name)	
1	-98.94764

○ año_cierre y hora_cierre

```

select [codigo_cierre], [fecha_cierre], [año_cierre], [mes_cierre], [hora_cierre], [delegacion_cierre]
from [test].[dbo].[incidentevial2dsem2020]
where [año_cierre] = (select min([año_cierre]) from [test].[dbo].[incidentevial2dsem2020])

```

	codigo_cierre	fecha_cierre	año_cierre	mes_cierre	hora_cierre	delegacion_cierre
1	(A) La unidad de atención a emergencias fue desp...	2020-06-01 00:00:00.0000000 -05:00	2020	Junio	12:31:56	ALVARO OBREGON
2	(A) La unidad de atención a emergencias fue desp...	2020-06-01 00:00:00.0000000 -05:00	2020	Junio	23:00:41	MIGUEL HIDALGO
3	(D) El incidente reportado se registró en dos o más ...	2020-06-02 00:00:00.0000000 -05:00	2020	Junio	11:26:06	ALVARO OBREGON
4	(D) El incidente reportado se registró en dos o más ...	2020-06-02 00:00:00.0000000 -05:00	2020	Junio	12:06:03	ALVARO OBREGON
5	(D) El incidente reportado se registró en dos o más ...	2020-06-02 00:00:00.0000000 -05:00	2020	Junio	12:56:54	ALVARO OBREGON
6	(D) El incidente reportado se registró en dos o más ...	2020-06-03 00:00:00.0000000 -05:00	2020	Junio	18:55:29	MIGUEL HIDALGO
7	(D) El incidente reportado se registró en dos o más ...	2020-06-03 00:00:00.0000000 -05:00	2020	Junio	18:57:29	MIGUEL HIDALGO
8	(D) El incidente reportado se registró en dos o más ...	2020-06-03 00:00:00.0000000 -05:00	2020	Junio	18:59:57	MIGUEL HIDALGO
9	(D) El incidente reportado se registró en dos o más ...	2020-06-03 00:00:00.0000000 -05:00	2020	Junio	19:45:50	MIGUEL HIDALGO
10	(D) El incidente reportado se registró en dos o más ...	2020-06-03 00:00:00.0000000 -05:00	2020	Junio	21:13:09	CUAJIMALPA
11	(A) La unidad de atención a emergencias fue desp...	2020-06-04 00:00:00.0000000 -05:00	2020	Junio	20:10:12	MAGDALENA CO...
12	(A) La unidad de atención a emergencias fue desp...	2020-06-04 00:00:00.0000000 -05:00	2020	Junio	19:23:17	MAGDALENA CO...
13	(A) La unidad de atención a emergencias fue desp...	2020-06-05 00:00:00.0000000 -05:00	2020	Junio	05:15:20	ALVARO OBREGON
14	(A) La unidad de atención a emergencias fue desp...	2020-06-05 00:00:00.0000000 -05:00	2020	Junio	16:33:26	MIGUEL HIDALGO
15	(N) La unidad de atención a emergencias fue desp...	2020-06-05 00:00:00.0000000 -05:00	2020	Junio	23:57:31	AZCAPOTZALCO
16	(N) La unidad de atención a emergencias fue desp...	2020-06-06 00:00:00.0000000 -05:00	2020	Junio	00:37:05	ALVARO OBREGON
17	(A) La unidad de atención a emergencias fue desp...	2020-06-06 00:00:00.0000000 -05:00	2020	Junio	03:59:04	ALVARO OBREGON
18	(A) La unidad de atención a emergencias fue desp...	2020-06-07 00:00:00.0000000 -05:00	2020	Junio	01:38:00	MIGUEL HIDALGO
19	(N) La unidad de atención a emergencias fue desp...	2020-06-07 00:00:00.0000000 -05:00	2020	Junio	23:31:17	MAGDALENA CO...
20	(D) El incidente reportado se registró en dos o más ...	2020-06-07 00:00:00.0000000 -05:00	2020	Junio	23:44:44	CUAJIMALPA
21	(A) La unidad de atención a emergencias fue desp...	2020-06-08 00:00:00.0000000 -05:00	2020	Junio	10:05:24	ALVARO OBREGON

```

select [codigo_cierre], [fecha_cierre], [año_cierre], [mes_cierre], [hora_cierre], [delegacion_cierre]
from [test].[dbo].[incidentevial2dsem2020]
where [año_cierre] = (select max([año_cierre]) from [test].[dbo].[incidentevial2dsem2020])

```

	codigo_cierre	fecha_cierre	año_cierre	mes_cierre	hora_cierre	delegacion_cierre
1	(A) La unidad de atención a emergencias fue desp...	2020-06-01 00:00:00.0000000 -05:00	2020	Junio	12:31:56	ALVARO OBREGON
2	(A) La unidad de atención a emergencias fue desp...	2020-06-01 00:00:00.0000000 -05:00	2020	Junio	23:00:41	MIGUEL HIDALGO
3	(D) El incidente reportado se registró en dos o más ...	2020-06-02 00:00:00.0000000 -05:00	2020	Junio	11:26:06	ALVARO OBREGON
4	(D) El incidente reportado se registró en dos o más ...	2020-06-02 00:00:00.0000000 -05:00	2020	Junio	12:06:03	ALVARO OBREGON
5	(D) El incidente reportado se registró en dos o más ...	2020-06-02 00:00:00.0000000 -05:00	2020	Junio	12:56:54	ALVARO OBREGON
6	(D) El incidente reportado se registró en dos o más ...	2020-06-03 00:00:00.0000000 -05:00	2020	Junio	18:55:29	MIGUEL HIDALGO
7	(D) El incidente reportado se registró en dos o más ...	2020-06-03 00:00:00.0000000 -05:00	2020	Junio	18:57:29	MIGUEL HIDALGO
8	(D) El incidente reportado se registró en dos o más ...	2020-06-03 00:00:00.0000000 -05:00	2020	Junio	18:59:57	MIGUEL HIDALGO
9	(D) El incidente reportado se registró en dos o más ...	2020-06-03 00:00:00.0000000 -05:00	2020	Junio	19:45:50	MIGUEL HIDALGO
10	(D) El incidente reportado se registró en dos o más ...	2020-06-03 00:00:00.0000000 -05:00	2020	Junio	21:13:09	CUAJIMALPA
11	(A) La unidad de atención a emergencias fue desp...	2020-06-04 00:00:00.0000000 -05:00	2020	Junio	20:10:12	MAGDALENA CO...
12	(A) La unidad de atención a emergencias fue desp...	2020-06-04 00:00:00.0000000 -05:00	2020	Junio	19:23:17	MAGDALENA CO...
13	(A) La unidad de atención a emergencias fue desp...	2020-06-05 00:00:00.0000000 -05:00	2020	Junio	05:15:20	ALVARO OBREGON
14	(A) La unidad de atención a emergencias fue desp...	2020-06-05 00:00:00.0000000 -05:00	2020	Junio	16:33:26	MIGUEL HIDALGO
15	(N) La unidad de atención a emergencias fue desp...	2020-06-05 00:00:00.0000000 -05:00	2020	Junio	23:57:31	AZCAPOTZALCO
16	(N) La unidad de atención a emergencias fue desp...	2020-06-06 00:00:00.0000000 -05:00	2020	Junio	00:37:05	ALVARO OBREGON
17	(A) La unidad de atención a emergencias fue desp...	2020-06-06 00:00:00.0000000 -05:00	2020	Junio	03:59:04	ALVARO OBREGON
18	(A) La unidad de atención a emergencias fue desp...	2020-06-07 00:00:00.0000000 -05:00	2020	Junio	01:38:00	MIGUEL HIDALGO
19	(N) La unidad de atención a emergencias fue desp...	2020-06-07 00:00:00.0000000 -05:00	2020	Junio	23:31:17	MAGDALENA CO...
20	(D) El incidente reportado se registró en dos o más ...	2020-06-07 00:00:00.0000000 -05:00	2020	Junio	23:44:44	CUAJIMALPA
21	(A) La unidad de atención a emergencias fue desp...	2020-06-08 00:00:00.0000000 -05:00	2020	Junio	10:05:24	ALVARO OBREGON

Se encontraron los rangos de las siguientes columnas, los cuales se explicarán punto por punto el significado de cada una.

- **Incidente:** Esta consulta fue un poco difícil aplicar ‘distinct’, ya que nos marcaba un error el cual no era posible comparar los datos por el tipo de dato que presentaban.

```
select distinct [incidente_c4] from [test].[dbo].[incidentevial2dsem2020]
```

100 %

Messages

Msg 421, Level 16, State 1, Line 21
The text data type cannot be selected as DISTINCT because it is not comparable.

Completion time: 2021-08-30T17:48:17.9489543-05:00

Los siguientes valores que toman la columna están explicados en cada uno, el cual describe el tipo de accidente registrado en la columna, así como si hubo lesiones y que tipo de transporte se utilizaba.

```
select [incidente_c4] from [test].[dbo].[incidentevial2dsem2020]
```

100 %

Results Messages

	incidente_c4
1	accidente-choque con lesionados
2	accidente-choque con lesionados
3	accidente-choque sin lesionados
4	accidente-choque sin lesionados
5	accidente-choque sin lesionados
6	accidente-motociclista
7	accidente-motociclista
8	accidente-motociclista
9	accidente-choque sin lesionados
10	accidente-choque sin lesionados

- **tipo_entrada:** La columna nos ilustra el tipo de denuncia que se presentó al accidente vial reportado, si fue mediante un botón de auxilio, Llamada directa al 911, radio o si se detectó en las cámaras de la ciudad.

```
select distinct [tipo_entrada] from [test].[dbo].[incidentevial2dsem2020]
```

100 %

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:** Se reporta la clase de alarma utilizada durante el incidente, si fue una emergencia, un posible delito o atentado, falsa alarma o si se hizo uso de urgencias médicas en el lugar.

```
select distinct [clas_con_f_alarma] from [test].[dbo].[incidentevial2dsem2020]
```

100 %

Results Messages

	clas_con_f_alarma
1	EMERGENCIA
2	DELITO
3	FALSA ALARMA
4	URGENCIAS MEDICAS

- **Delegación:** Podemos tomar como referencias las alcaldías en las cuales se reportan más casos de accidentes viales, ya que se tendrían que tomar medidas para ello, por eso es importante el saber el conteo de cuantos accidentes por delegaciones existe.

```
select distinct [delegacion_inicio] from [test].[dbo].[incidentevial2dsem2020]
```

100 %

Results Messages

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

Se contabilizo la cantidad de nulos existentes en cada uno de los puntos anteriores:

- **Incidentes_C4:** En esta consulta también tuvimos problemas al verificar si existían nulos.

```
select count([incidente_c4]) from [test].[dbo].[incidentevial2dsem2020] where [incidente_c4] = 'NULL'
```

100 %

Messages

Msg 402, Level 16, State 1, Line 28
The data types text and varchar are incompatible in the equal to operator.

Completion time: 2021-08-30T18:05:19.0596104-05:00

- **Tipo entrada:**

```
select count([tipo_entrada]) from [test].[dbo].[incidentevial2dsem2020] where [tipo_entrada] = 'NULL'
```

100 %

Results Messages

	(No column name)
1	0

- **clas_con_f_alarma:**

```
select count([clas_con_f_alarma]) from [test].[dbo].[incidentevial2dsem2020] where [clas_con_f_alarma] = 'NULL'
```

100 %

Results Messages

	(No column name)
1	0

- **Delegación:**

```
select count([delegacion_inicio]) from [test].[dbo].[incidentevial2dsem2020] where [delegacion_inicio] = 'NULL'
```

100 %

Results Messages

	(No column name)
1	2

Al hacer nuestro análisis con los datos de la base de accidentes viales, nos dimos cuenta que nos daba el error al intentar buscar los nulos en los Incidentes c4, por lo que solamente encontramos la descripción del reporte, pero no pudimos comprobar si había anomalías. Como segundo punto encontramos nulos en la parte de Delegaciones algo que se nos hizo bastante interesante porque tal vez al meter los datos se tuvieron algunos problemas, algunas suposiciones obvias que se podrían hacer es que la persona que ingreso el reporte no lleno de manera correcta los campos, como segunda opción puede ser la réplica de los datos , lo que ocasionaría que los datos se repitan pero sin llenar todos los campos, para saber qué fue lo que pudo pasar se tendrían que analizar las fechas de los incidentes y si ay algún registro del error o checar las fechas en las que fue levantado el reporte para darnos una idea del por qué los valores que no concuerdan con los demás o la columna en sí.

Conclusión

En la presente practica se pude hacer un repaso de la estructura de las consultas que se utilizan para recabar información de una base de datos. Además, se analizaron los datos con los que tratamos de responder algunos puntos de las practicas, cuáles eran los significados y los valores que podrían llegar a tomar cada uno de ellos; en algunas columnas encontramos algunas inconsistencias de los calores que tomaban, ya que había algunos nulos, por lo que se puede conjeturar que tal vez al subir el reporte del accidente vial, se tuvo un error de dedos del cual no se llenó ese campo, otro caso posible es que se pudieron llegar a duplicar los datos pero sin todos los datos completos, por lo cual se tendría que ver la fecha y hora del caso, para un análisis más profundo de lo que pudo ocasionar ese espacio vacío.

Es interesante analizar este tipo de bases de datos ya que con esto podemos plantear soluciones para diferentes problemas, ya que con la información obtenida supongamos que podemos realizar una búsqueda de las colonias con más accidentes viales y su zona, por lo que con esto se tomaran las medidas necesarias y se podrá ver cuáles son los accidentes más comunes en el área, así se podrá evitar o disminuir los reportes que llegan día a día sobre estos casos.