

INSTITUTO POLITÉCNICO NACIONAL ESCUELA SUPERIOR DE CÓMPUTO



DATA MINING

Práctica 1. Importación

Graciano Herrera Gabriel
Meza Zamora Abraham Manuel

15 de marzo de 2022

Índice

1. Introducción	2
2. Desarrollo	3
2.1. Descargar el dataset de incidentes viales	3
2.2. Exportar el archivo CSV en el manejador de base de datos seleccionado	3
2.3. Indicar el número de registros del dataset en el manejador	5
2.4. Rango de los campos relacionados (valor mínimo y máximo)	5
2.4.1. Fecha - Mínimo	5
2.4.2. Fecha - Máximo	5
2.4.3. Latitud - Mínimo	6
2.4.4. Latitud - Máximo	6
2.4.5. Longitud - Mínimo	7
2.4.6. Longitud - Máximo	7
2.4.7. Año cierre - Mínimo	8
2.4.8. Año cierre - Máximo	8
2.4.9. Hora cierre - Mínimo	9
2.4.10. Hora cierre - Máxima	9
2.5. ¿Cuáles son los valores que toman las siguientes columnas (rango, i.e. valores posibles no repetidos) y que significado tiene	10
2.5.1. incidente_c4	10
2.5.2. tipo_entrada	11
2.5.3. clas_con_f_alarma	11
2.5.4. delegacion	11
2.6. Contar la cantidad de NULL o NULOS encontrados en las 4 columnas anteriores	14
2.7. 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	14
3. Conclusiones	14
3.1. Gabriel Graciano Herrera	14
3.2. Meza Zamora Abraham Manuel	15

1. Introducción

Las siglas CSV significan Comma Separated Values (valores separados por comas, en inglés) y un archivo CSV es cualquier archivo de texto en el cual los valores están separados por comas, creando así, una especie de tabla en filas y columnas donde las columnas se delimitan por cada punto y coma (;), mientras que cada fila se define mediante una línea adicional en el texto. Es por esto por lo que los archivos .csv están asociados directamente a la creación de tablas de contenido. [1]

Comúnmente, los archivos CSV se leen desde Excel para su posterior manipulación, sin embargo, es posible también importar estos datos a una base de datos y trabajar con ellos como con cualquier otra base de datos.

En el desarrollo de esta práctica, se hace uso de SQL server para manipular los datos de un archivo CSV. Microsoft SQL Server es un sistema de gestión de base de datos relacional desarrollado como un servidor que da servicio a otras aplicaciones de software que pueden funcionar ya sea en la misma computadora o en otra a través de una red. [2]

Para la manipulación de estos datos, es necesario tener conocimientos básicos de SQL. El Lenguaje de Consulta Estructurado popularmente conocido por sus siglas en inglés como SQL, es un tipo de lenguaje de programación que ayuda a solucionar problemas específicos o relacionados con la definición, manipulación e integridad de la información representada por los datos que se almacenan en las bases de datos. [3]

SQL posee comandos referentes a:

- Un lenguaje de definición de datos o DDL en inglés, que permite:
 - La creación de la estructura o esquema de base de datos.
 - La modificación de dicha estructura.
- Un lenguaje de manipulación de datos o DML en inglés, que hace posible:
 - La inserción de datos en tablas.
 - Las consultas sobre los datos de estas tablas.
 - La actualización de los datos contenidos en estas tablas.
 - La eliminación de los registros de estas tablas.
- Peticiones de información más complejas que incluyen JOINS y subconsultas.
- Integridad: el DDL incluye también comandos para especificar las restricciones de integridad que el DBMS debe hacer cumplir.
- Definición de vistas o tablas derivadas.
- Funciones de agrupamiento: que permiten hacer cálculos de resúmenes.
- Control de Transacciones: como unidad de trabajo lógica, unidad de recuperación y mecanismo de concurrencia.
- Autorización: incluye comandos para otorgar los privilegios de acceso a las tablas, vistas y otros elementos de base de datos.

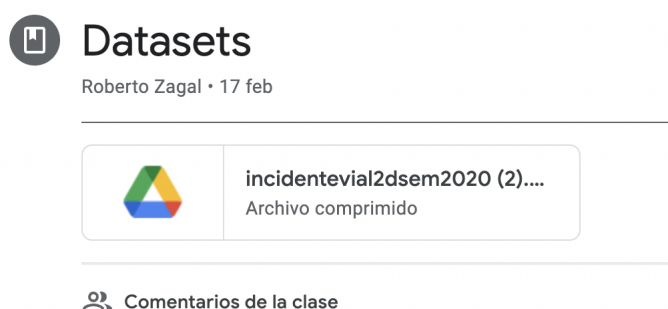
- SQL incorporado y dinámico: esto quiere decir que se puede incorporar comandos SQL en lenguajes de programación como C++, PHP, Java, etc.

Para esta práctica se utilizó un dataset con los incidentes viales de la CDMX en el último semestre del 2020, el dataset se encontraba en un archivo CSV" por lo que se tuvo que realizar una interpretación de los datos previamente a la importación para darles un tipo de dato correcto en la BD. Además, se utilizó el gestor de BD Mysql con una BD de manera local.

2. Desarrollo

2.1. Descargar el dataset de incidentes viales

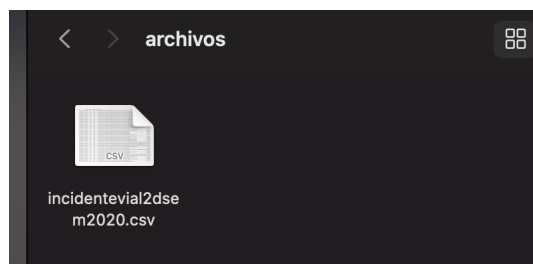
1. En la página de classroom de la clase damos clic sobre el enlace de "Datasets".



2. Posteriormente en la esquina superior derecha damos clic en el botón de descargar.



3. Finalmente descomprimos y obtenemos el siguiente archivo con extensión .csv .



2.2. Exportar el archivo CSV en el manejador de base de datos seleccionado

1. Creamos la tabla en la cual vamos a meter los datos del csv

```
CREATE TABLE incidencias (
    folio TEXT,
    fecha_creacion TEXT,
    hora_creacion TEXT,
    dia_semana TEXT,
    codigo_cierre TEXT,
    fecha_cierre TEXT,
    ano_cierre INT,
    mes_cierre TEXT,
    hora_cierre TEXT,
    delegacion_inicio TEXT,
    incidente_c4 TEXT,
    latitud DOUBLE,
    longitud DOUBLE,
    clas_con_f_alarma TEXT,
    tipo_entrada TEXT,
    delegacion_cierre TEXT,
    geopoint TEXT,
    mes INT
);
```

2. Exportamos los datos con la siguiente instrucción

```
LOAD DATA INFILE 'archivos/incidentevial2dsem2020.csv'
INTO TABLE incidencias
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS;
```

3. Verificamos que la tabla haya sido importada con éxito con el siguiente comando

```
SELECT * FROM incidencias;
```

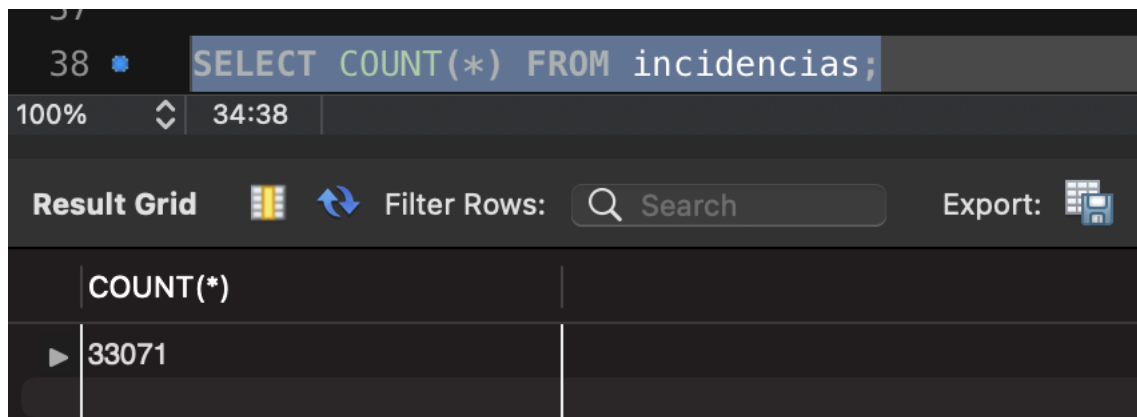
folio	fecha_creaci...	hora_creacion	dia_semana	codigo_cierre	fecha_cierre	ano_cierre	mes_cierre	hora_cierre	hora_cier
C5/200630/04173	30/06/2020	15:06:18	Martes	(A) La unidad de atención a emergencias fue de...	2020-06-30 00:00:00.0000000	2020	Junio	16:21:46	
C5/200605/07195	05/06/2020	19:46:13	Viernes	(N) La unidad de atención a emergencias fue d...	2020-06-05 00:00:00.0000000	2020	Junio	19:54:15	
C5/200630/04306	30/06/2020	15:22:52	Martes	(D) El incidente reportado se registró en dos o...	2020-06-30 00:00:00.0000000	2020	Junio	15:25:47	
IZ/200605/00026	05/06/2020	00:02:58	Viernes	(A) La unidad de atención a emergencias fue de...	2020-06-05 00:00:00.0000000	2020	Junio	00:31:21	
C5/200630/04376	30/06/2020	15:30:02	Martes	(D) El incidente reportado se registró en dos o...	2020-06-30 00:00:00.0000000	2020	Junio	16:03:11	
IZ/200605/02179	05/06/2020	08:43:41	Viernes	(A) La unidad de atención a emergencias fue de...	2020-06-05 00:00:00.0000000	2020	Junio	09:32:02	
C5/200630/04344	30/06/2020	15:23:20	Martes	(A) La unidad de atención a emergencias fue de...	2020-06-30 00:00:00.0000000	2020	Junio	19:02:18	
AO/200605/08962	05/06/2020	23:02:58	Viernes	(N) La unidad de atención a emergencias fue de...	2020-06-05 00:00:00.0000000	2020	Junio	23:57:31	
C5/200630/04438	30/06/2020	15:36:20	Martes	(A) La unidad de atención a emergencias fue de...	2020-06-30 00:00:00.0000000	2020	Junio	16:55:21	
C5/200604/04561	04/06/2020	16:29:04	Jueves	(A) La unidad de atención a emergencias fue de...	2020-06-05 00:00:00.0000000	2020	Junio	00:13:49	
C5/200630/04454	30/06/2020	15:40:49	Martes	(A) La unidad de atención a emergencias fue de...	2020-06-30 00:00:00.0000000	2020	Junio	17:34:07	
C5/200604/04606	04/06/2020	16:34:12	Jueves	(A) La unidad de atención a emergencias fue de...	2020-06-05 00:00:00.0000000	2020	Junio	17:32:58	
C5/200630/04472	30/06/2020	15:41:14	Martes	(N) La unidad de atención a emergencias fue d...	2020-06-30 00:00:00.0000000	2020	Junio	15:59:17	
C5/200605/00428	05/06/2020	00:58:28	Viernes	(A) La unidad de atención a emergencias fue de...	2020-06-05 00:00:00.0000000	2020	Junio	04:54:08	
C5/200630/04485	30/06/2020	15:43:06	Martes	(D) El incidente reportado se registró en dos o...	2020-06-30 00:00:00.0000000	2020	Junio	15:49:08	
C5/200605/00476	05/06/2020	01:07:10	Viernes	(A) La unidad de atención a emergencias fue de...	2020-06-05 00:00:00.0000000	2020	Junio	01:53:22	
C5/200630/04590	30/06/2020	15:55:56	Martes	(N) La unidad de atención a emergencias fue despachada, llegó al lugar de los hechos, pero en el sitio del evento nadie solicitó el ap...	2020-06-30 00:00:00.0000000	2020	Junio	18:51:21	
C5/200605/00741	05/06/2020	01:53:56	Viernes	(A) La unidad de atención a emergencias fue de...	2020-06-05 00:00:00.0000000	2020	Junio	05:22:19	
C5/200630/04599	30/06/2020	15:56:20	Martes	(A) La unidad de atención a emergencias fue de...	2020-06-30 00:00:00.0000000	2020	Junio	18:51:21	
C5/200605/01923	05/06/2020	07:50:05	Viernes	(A) La unidad de atención a emergencias fue de...	2020-06-05 00:00:00.0000000	2020	Junio	08:41:13	

Figura 1: Podemos ver que la tabla se ha importado de manera exitosa.

2.3. Indicar el número de registros del dataset en el manejador

Para indicar el número de registros ejecutamos la siguiente consulta

```
SELECT COUNT(*) FROM incidencias;
```



COUNT(*)
33071

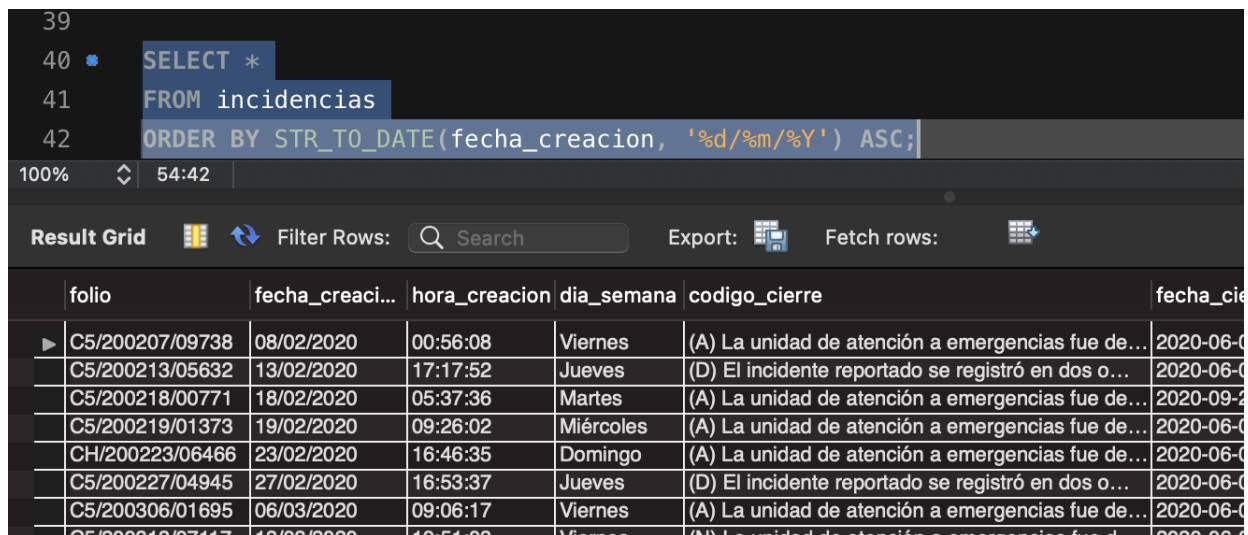
Figura 2: El número de registros es de 33071.

2.4. Rango de los campos relacionados (valor mínimo y máximo)

2.4.1. Fecha - Mínimo

Ingresamos la siguiente consulta en el manejador y ejecutamos

```
SELECT *  
FROM incidencias  
ORDER BY STR_TO_DATE( fecha_creacion , '%d/%m/%Y' ) ASC;
```



	folio	fecha_creaci...	hora_creacion	dia_semana	codigo_cierre	fecha_cie
▶	C5/200207/09738	08/02/2020	00:56:08	Viernes	(A) La unidad de atención a emergencias fue de...	2020-06-0
	C5/200213/05632	13/02/2020	17:17:52	Jueves	(D) El incidente reportado se registró en dos o...	2020-06-0
	C5/200218/00771	18/02/2020	05:37:36	Martes	(A) La unidad de atención a emergencias fue de...	2020-09-2
	C5/200219/01373	19/02/2020	09:26:02	Miércoles	(A) La unidad de atención a emergencias fue de...	2020-06-0
	CH/200223/06466	23/02/2020	16:46:35	Domingo	(A) La unidad de atención a emergencias fue de...	2020-06-0
	C5/200227/04945	27/02/2020	16:53:37	Jueves	(D) El incidente reportado se registró en dos o...	2020-06-0
	C5/200306/01695	06/03/2020	09:06:17	Viernes	(A) La unidad de atención a emergencias fue de...	2020-06-0
	C5/200312/07117	12/03/2020	10:51:02	Viernes	(A) La unidad de atención a emergencias fue d...	2020-06-0

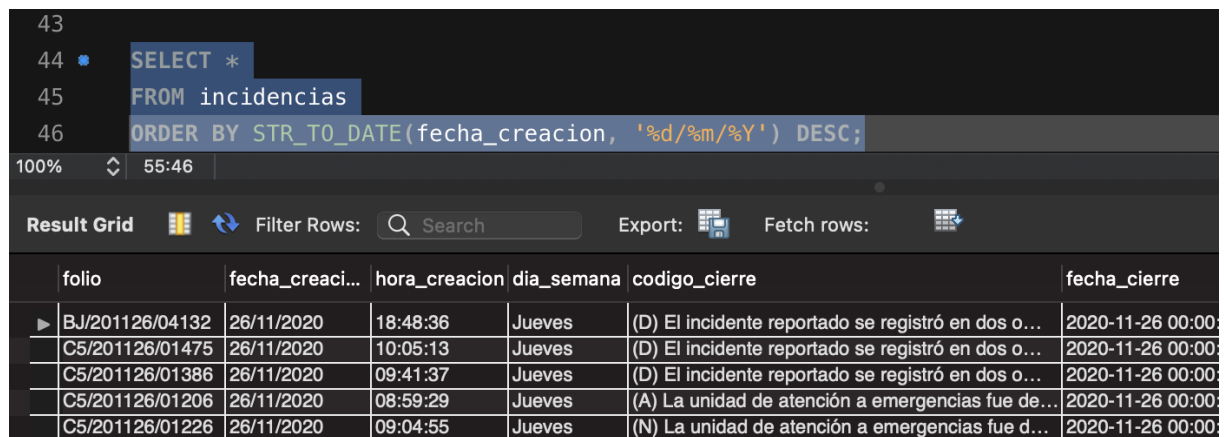
Figura 3: Vemos que la fecha más antigua es del 8 de Febrero de 2020.

2.4.2. Fecha - Máximo

Ingresamos la siguiente consulta en el manejador y ejecutamos

```
SELECT *  
FROM incidencias
```

```
ORDER BY STR_TO_DATE( fecha_creacion , '%d/%m/%Y' ) DESC;
```



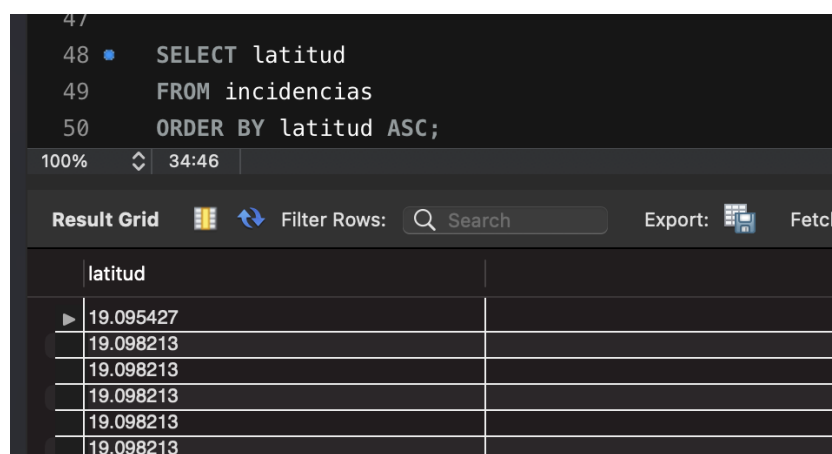
	folio	fecha_creaci...	hora_creacion	dia_semana	codigo_cierre	fecha_cierre
▶	BJ/201126/04132	26/11/2020	18:48:36	Jueves	(D) El incidente reportado se registró en dos o...	2020-11-26 00:00:
	C5/201126/01475	26/11/2020	10:05:13	Jueves	(D) El incidente reportado se registró en dos o...	2020-11-26 00:00:
	C5/201126/01386	26/11/2020	09:41:37	Jueves	(D) El incidente reportado se registró en dos o...	2020-11-26 00:00:
	C5/201126/01206	26/11/2020	08:59:29	Jueves	(A) La unidad de atención a emergencias fue de...	2020-11-26 00:00:
	C5/201126/01226	26/11/2020	09:04:55	Jueves	(N) La unidad de atención a emergencias fue d...	2020-11-26 00:00:

Figura 4: Vemos que la fecha más reciente es del 26 de Noviembre de 2020.

2.4.3. Latitud - Mínimo

Ingresamos la siguiente consulta en el manejador y ejecutamos

```
SELECT latitud  
FROM incidencias  
ORDER BY latitud ASC;
```



	latitud
▶	19.095427
	19.098213
	19.098213
	19.098213
	19.098213
	19.098213

Figura 5: Vemos que la latitud mínima es de 19.095427.

2.4.4. Latitud - Máximo

Ingresamos la siguiente consulta en el manejador y ejecutamos

```
SELECT latitud  
FROM incidencias  
ORDER BY latitud DESC;
```

```

51
52 • SELECT latitud
53 FROM incidencias
54 ORDER BY latitud DESC;
100% 15:52
Result Grid Filter Rows: Search Export
latitud
19.57671
19.57534
19.57514
19.574963
19.57109
19.57109
19.57109

```

Figura 6: Vemos que la latitud máxima es de 19.57671.

2.4.5. Longitud - Mínimo

Ingresamos la siguiente consulta en el manejador y ejecutamos

```

SELECT longitud
FROM incidencias
ORDER BY longitud ASC;

```

```

55
56 • SELECT longitud
57 FROM incidencias
58 ORDER BY longitud ASC;
100% 23:58
Result Grid Filter Rows: Search Export
longitud
-99.348434
-99.348434
-99.348434

```

Figura 7: Vemos que la longitud mínima es de -99.348434.

2.4.6. Longitud - Máximo

Ingresamos la siguiente consulta en el manejador y ejecutamos

```

SELECT longitud
FROM incidencias
ORDER BY longitud DESC;

```

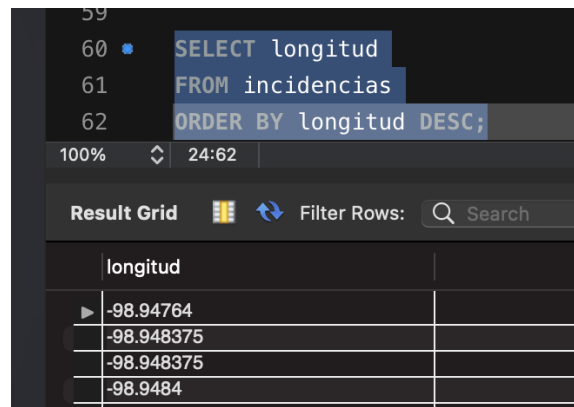



Figura 8: Vemos que la longitud máxima es de -98.94764.

2.4.7. Año cierre - Mínimo

Ingresamos la siguiente consulta en el manejador y ejecutamos

```

SELECT ano_cierre
FROM incidencias
ORDER BY ano_cierre ASC;

```

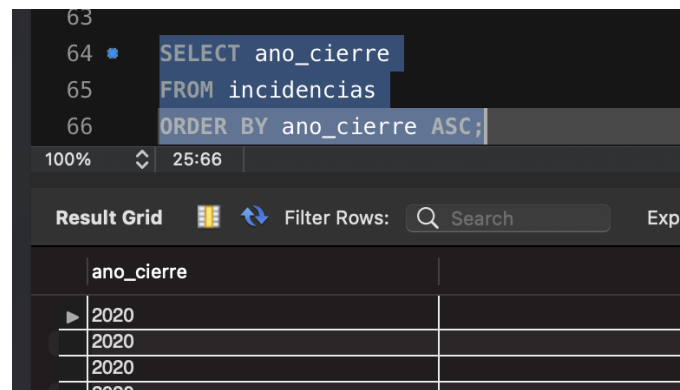


Figura 9: Vemos que el año más antiguo es 2020.

2.4.8. Año cierre - Máximo

Ingresamos la siguiente consulta en el manejador y ejecutamos

```

SELECT ano_cierre
FROM incidencias
ORDER BY ano_cierre DESC;

```

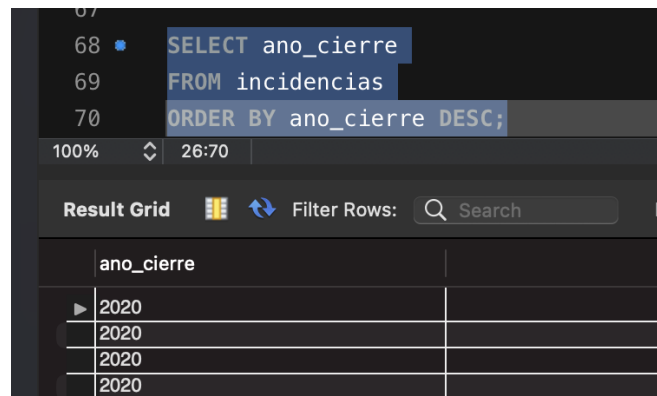


Figura 10: Vemos que el año más reciente es 2020.

2.4.9. Hora cierre - Mínimo

Ingresamos la siguiente consulta en el manejador y ejecutamos

```

SELECT hora_cierre
FROM incidencias
ORDER BY hora_cierre ASC;

```

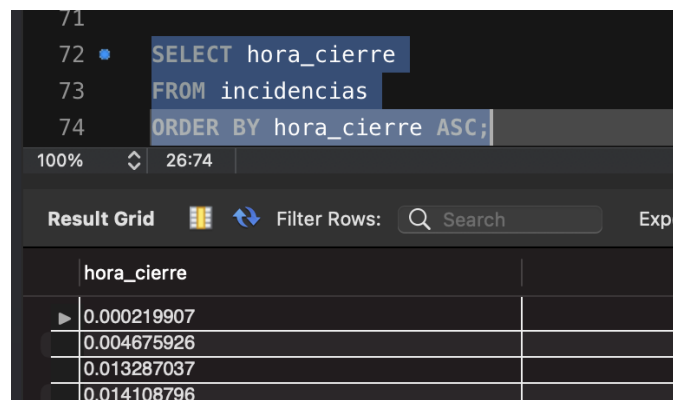


Figura 11: Vemos la hora más antigua es 0.000219907.

2.4.10. Hora cierre - Máxima

Ingresamos la siguiente consulta en el manejador y ejecutamos

```

SELECT hora_cierre
FROM incidencias
ORDER BY hora_cierre DESC;

```

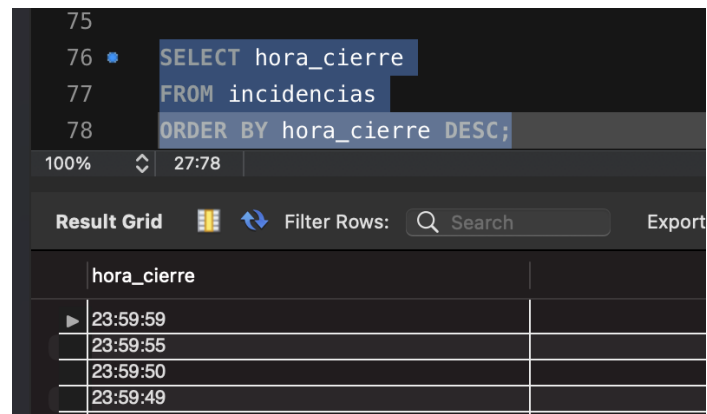


Figura 12: Vemos la hora más antigua es 23:59:59.

2.5. ¿Cuáles son los valores que toman las siguientes columnas (rango, i.e. valores posibles no repetidos) y que significado tiene

2.5.1. incidente_c4

Ingresamos la siguiente consulta en el manejador y ejecutamos

```
SELECT DISTINCT incidente_c4 FROM incidencias;
```

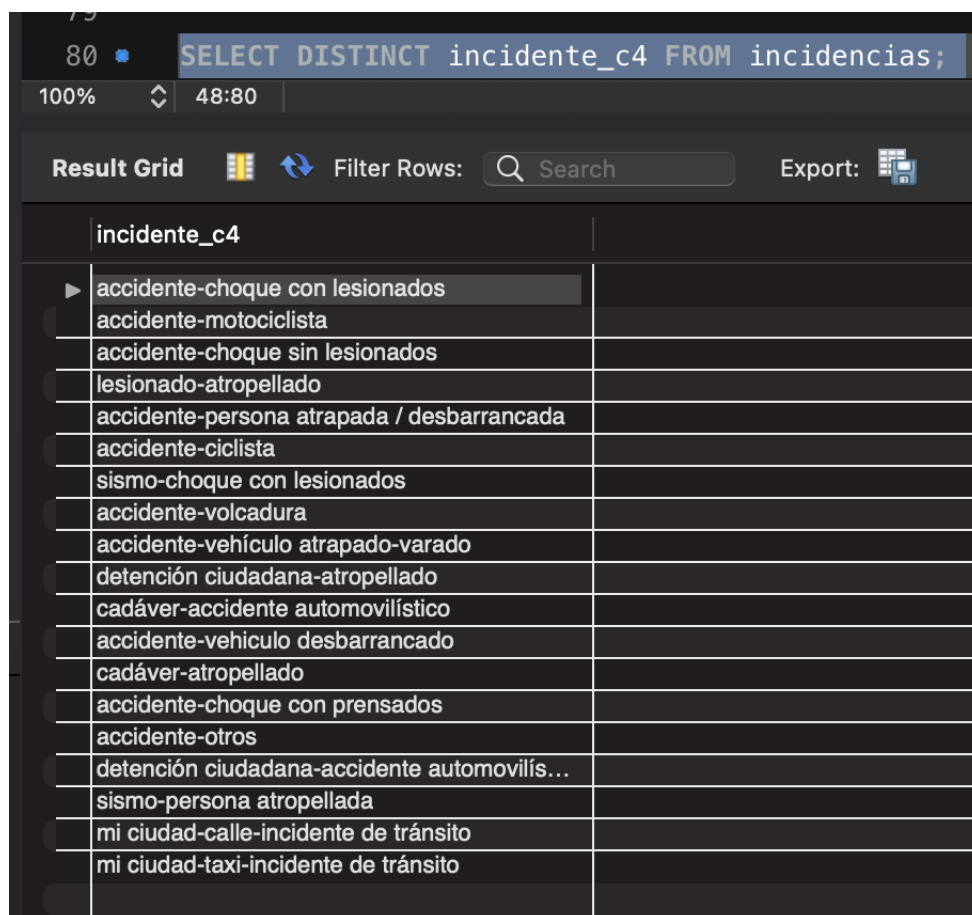
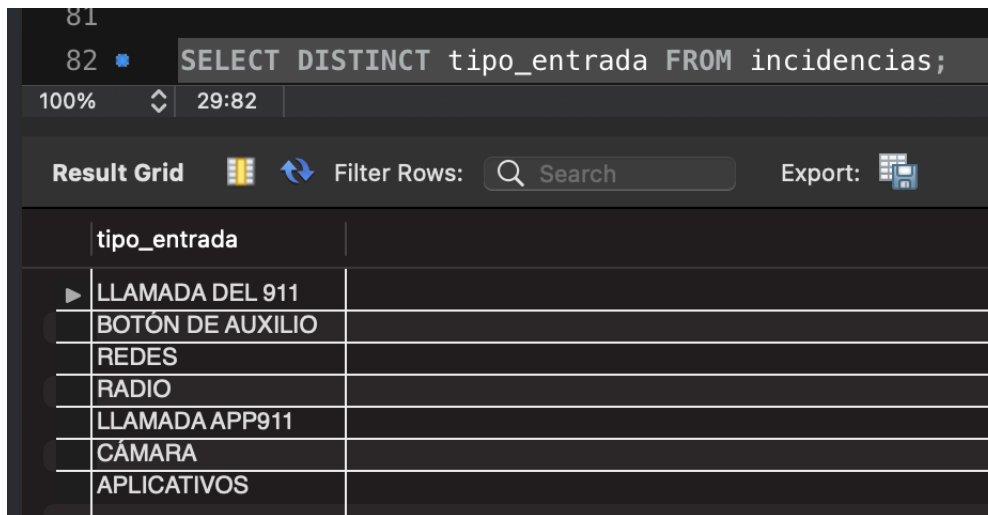


Figura 13: Vemos los distintos tipos de incidentes tipo C4.

2.5.2. tipo_entrada

Ingresamos la siguiente consulta en el manejador y ejecutamos

```
SELECT DISTINCT tipo_entrada FROM incidencias;
```



The screenshot shows a database query interface. At the top, a SQL query is entered: `SELECT DISTINCT tipo_entrada FROM incidencias;`. Below the query, the interface displays a 'Result Grid' with a table of results. The table has one column, 'tipo_entrada', and lists seven distinct values: 'LLAMADA DEL 911', 'BOTÓN DE AUXILIO', 'REDES', 'RADIO', 'LLAMADA APP911', 'CÁMARA', and 'APLICATIVOS'.

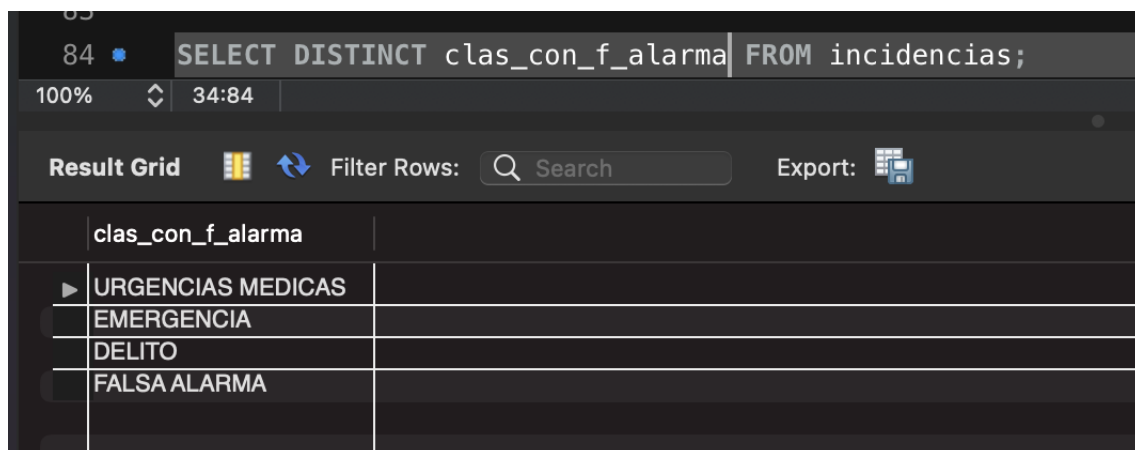
tipo_entrada
LLAMADA DEL 911
BOTÓN DE AUXILIO
REDES
RADIO
LLAMADA APP911
CÁMARA
APLICATIVOS

Figura 14: Vemos los distintos tipos de incidentes tipo entrada.

2.5.3. clas_con_f_alarma

Ingresamos la siguiente consulta en el manejador y ejecutamos

```
SELECT DISTINCT clas_con_f_alarma FROM incidencias;
```



The screenshot shows a database query interface. At the top, a SQL query is entered: `SELECT DISTINCT clas_con_f_alarma FROM incidencias;`. Below the query, the interface displays a 'Result Grid' with a table of results. The table has one column, 'clas_con_f_alarma', and lists four distinct values: 'URGENCIAS MEDICAS', 'EMERGENCIA', 'DELITO', and 'FALSA ALARMA'.

clas_con_f_alarma
URGENCIAS MEDICAS
EMERGENCIA
DELITO
FALSA ALARMA

Figura 15: Vemos los distintos tipos de clas con f alarma.

2.5.4. delegacion

Ingresamos la siguiente consulta en el manejador y ejecutamos

```
SELECT DISTINCT delegacion_inicio , delegacion_cierre FROM incidencias;
```

86 • `SELECT DISTINCT delegacion_inicio, delegacion_cierre FROM incidencias;`

100% 72:86

Result Grid Filter Rows: Search Export:

	delegacion_inicio	delegacion_cierre	
►	GUSTAVO A. MADERO	GUSTAVO A. MADERO	
●	ALVARO OBREGON	ALVARO OBREGON	
●	XOCHIMILCO	XOCHIMILCO	
●	TLALPAN	TLALPAN	
●	IZTAPALAPA	IZTAPALAPA	
●	AZCAPOTZALCO	AZCAPOTZALCO	
●	CUAUHTEMOC	CUAUHTEMOC	
●	MIGUEL HIDALGO	MIGUEL HIDALGO	
●	VENUSTIANO CARRANZA	VENUSTIANO CARRANZA	
●	BENITO JUAREZ	BENITO JUAREZ	
●	COYOACAN	ALVARO OBREGON	
●	COYOACAN	COYOACAN	
●	TLALPAN	COYOACAN	
●	IZTAPALAPA	IZTACALCO	
●	IZTACALCO	IZTACALCO	
●	IZTACALCO	IZTAPALAPA	
●	MAGDALENA CONTRERAS	MAGDALENA CONTRERAS	
●	TLAHUAC	TLAHUAC	
●	MILPA ALTA	MILPA ALTA	
●	IZTACALCO	VENUSTIANO CARRANZA	

Figura 16: Primera parte de la consulta de delegaciones.

	delegacion_inicio	delegacion_cierre	
●	CUAUHTEMOC	GUSTAVO A. MADERO	
●	BENITO JUAREZ	ALVARO OBREGON	
●	IZTAPALAPA	COYOACAN	
●	COYOACAN	TLALPAN	
●	CUAJIMALPA	CUAJIMALPA	
●	AZCAPOTZALCO	GUSTAVO A. MADERO	
●	MAGDALENA CONTRERAS	TLALPAN	
●	CUAUHTEMOC	MIGUEL HIDALGO	
●	XOCHIMILCO	TLAHUAC	
●	CUAUHTEMOC	AZCAPOTZALCO	
●	VENUSTIANO CARRANZA	GUSTAVO A. MADERO	
●	IZTACALCO	CUAUHTEMOC	
●	XOCHIMILCO	TLALPAN	
●	TLAHUAC	IZTAPALAPA	
●	ALVARO OBREGON	MAGDALENA CONTRERAS	
●	XOCHIMILCO	IZTAPALAPA	
●	MIGUEL HIDALGO	CUAUHTEMOC	
●	IZTAPALAPA	TLAHUAC	
●	VENUSTIANO CARRANZA	CUAUHTEMOC	
●	IZTAPALAPA	XOCHIMILCO	

Figura 17: Segunda parte de la consulta de delegaciones.

	delegacion_inicio	delegacion_cierre	
	CUAJIMALPA	ALVARO OBREGON	
	ALVARO OBREGON	CUAJIMALPA	
	GUSTAVO A. MADERO	AZCAPOTZALCO	
	AZCAPOTZALCO	CUAUHTEMOC	
	MIGUEL HIDALGO	ALVARO OBREGON	
	VENUSTIANO CARRANZA	IZTACALCO	
	ALVARO OBREGON	BENITO JUAREZ	
	XOCHIMILCO	MILPA ALTA	
	ALVARO OBREGON	MIGUEL HIDALGO	
	TLAHUAC	XOCHIMILCO	
	TLALPAN	XOCHIMILCO	
	GUSTAVO A. MADERO	VENUSTIANO CARRANZA	
	GUSTAVO A. MADERO	CUAUHTEMOC	
	BENITO JUAREZ	IZTACALCO	
	CUAUHTEMOC	BENITO JUAREZ	
	COYOACAN	IZTAPALAPA	
	BENITO JUAREZ	CUAUHTEMOC	
	ALVARO OBREGON	COYOACAN	
	IZTACALCO	GUSTAVO A. MADERO	
	CUAUHTEMOC	IZTACALCO	

Figura 18: Tercera parte de la consulta de delegaciones.

	delegacion_inicio	delegacion_cierre	
	CUAUHTEMOC	BENITO JUAREZ	
	COYOACAN	IZTAPALAPA	
	BENITO JUAREZ	CUAUHTEMOC	
	ALVARO OBREGON	COYOACAN	
	IZTACALCO	GUSTAVO A. MADERO	
	CUAUHTEMOC	IZTACALCO	
	BENITO JUAREZ	MIGUEL HIDALGO	
	BENITO JUAREZ	COYOACAN	
	CUAUHTEMOC	VENUSTIANO CARRANZA	
	MIGUEL HIDALGO	CUAJIMALPA	
	MIGUEL HIDALGO	BENITO JUAREZ	
	IZTACALCO	BENITO JUAREZ	
	TLALPAN	MAGDALENA CONTRERAS	
	BENITO JUAREZ	IZTAPALAPA	
	IZTAPALAPA	BENITO JUAREZ	
	ALVARO OBREGON	TLALPAN	
	TLALPAN	ALVARO OBREGON	
	MAGDALENA CONTRERAS	ALVARO OBREGON	
	COYOACAN	BENITO JUAREZ	
	VENUSTIANO CARRANZA	IZTAPALAPA	
	MAGDALENA CONTRERAS	MIGUEL HIDALGO	

Figura 19: Cuarta parte de la consulta de delegaciones.

	delegacion_inicio	delegacion_cierre	
	BENITO JUAREZ	MIGUEL HIDALGO	
	BENITO JUAREZ	COYOACAN	
	CUAUHTEMOC	VENUSTIANO CARRANZA	
	MIGUEL HIDALGO	CUAJIMALPA	
	MIGUEL HIDALGO	BENITO JUAREZ	
	IZTACALCO	BENITO JUAREZ	
	TLALPAN	MAGDALENA CONTRERAS	
	BENITO JUAREZ	IZTAPALAPA	
	IZTAPALAPA	BENITO JUAREZ	
	ALVARO OBREGON	TLALPAN	
	TLALPAN	ALVARO OBREGON	
	MAGDALENA CONTRERAS	ALVARO OBREGON	
	COYOACAN	BENITO JUAREZ	
	VENUSTIANO CARRANZA	IZTAPALAPA	
	MAGDALENA CONTRERAS	MIGUEL HIDALGO	
	CUAJIMALPA	MIGUEL HIDALGO	
	NULL	NULL	
	MIGUEL HIDALGO	AZCAPOTZALCO	
	MILPA ALTA	GUSTAVO A. MADERO	
	CUAJIMALPA	MAGDALENA CONTRERAS	

Figura 20: Quinta parte de la consulta de delegaciones.

2.6. Contar la cantidad de NULL o NULOS encontrados en las 4 columnas anteriores

Como podemos ver en la figura 20 los únicos valores que encontramos con NULL son en esos dos campos.

2.7. 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

La única anomalía que percibimos durante este análisis fue en la columna de hora_cierre, ya que en algunas entradas el valor estaba expresado como un valor decimal, y no como el formato que tenían el resto de las entradas. Es importante destacar que estos valores inconsistentes no llegaban ni al 10 % de la muestra.

3. Conclusiones

3.1. Gabriel Graciano Herrera

La práctica nos ayudo a reforzar lo aprendido en clase respecto a la importación de los datos de nuestro dataset, al momento de realizar la importación me encontré con algunos problemas como el tipo de dato que tenia que tener cada columna, la longitud en algunos casos, la codificación de los mismos, etc, pero me pude dar cuenta de este proceso de importación de los datos ya que si se realiza mal este proceso por ende los demás procesos estarán incorrectos o tendrán errores.

3.2. Meza Zamora Abraham Manuel

Durante el desarrollo de esta práctica, comenzamos a familiarizarnos con las herramientas que nos serán útiles durante el resto del curso. Así mismo, aprendimos sobre la importancia de identificar aquellos datos que sólo generan ruido y no son útiles para nuestros fines analíticos, tales como los datos nulos o los datos que no aportan información alguna, como sucedió con la consulta a la hora_cierre.