



Unidad Regional Centro

División de Ciencias Exactas y Naturales

Departamento de Matemáticas

Maestría en Ciencia de Datos 

Introducción a la Ciencia de Datos y sus Metodologías

**Proyecto Final: Creación de base de datos, conexión y análisis  
utilizando herramientas de Python.**

Dr. Juan Pablo Soto Barrera

Melissa Reyes Paz

## Tabla de contenidos.

Introducción. ....	3
Objetivos. ....	3
Descripción de fuente de datos.....	4
1. Manejo de datos en MySQL Workbench.....	5
1.1 Creación de base de datos.....	5
1.2 Creación de una vista.....	11
1.3 Creación de un procedimiento almacenado.....	14
1.4 Creación de una función. ....	16
2. Conexión a la base de datos desde Jupyter Notebook. ....	20
3. Análisis de la consulta de datos. ....	21
4. Repositorio de GitHub. ....	23
Referencias. ....	23

## **Introducción.**

Un manejador de base de datos (DBMS) es un conjunto de programas, los cuales, tiene el objetivo de modificar, almacenar, eliminar y extraer la información que contenga una base de datos (DB).

Entre los lenguajes más utilizados en un DBMS, cabe destacar el Lenguaje de Manipulación de datos o Data Manipulation Language (DML) para la realización de consultas y manipulación de datos. Especialmente, se utiliza el SQL (Structured Query Language), el DML más utilizado para gestionar datos relacionales, así como el Data Definition Language (DDL), utilizado para definir estructuras y funciones en la realización de consultas.

Es preciso decir que el uso de manejadores de bases de datos es fundamental para la manipulación de datos ya que permiten la identificación de aquellos datos que cumplen con ciertas características de una forma eficiente. Gracias a la complejidad o el tamaño de ciertas estructuras, es común el uso de las vistas, procedimientos almacenados y funciones en el mismo manejador, lo cual hace más fácil el manejo de la información.

Es posible generar una conexión con los manejadores de bases de datos desde un entorno de desarrollo integrado como Jupyter y RStudio. Esto resulta ser genial ya que ayuda a acelerar el proceso de consulta, extracción, manipulación y visualización de los datos.

Es por ello que en el presente documento se plasma el proceso de creación y manipulación de una base de datos utilizando MySQL Workbench y, posteriormente, la conexión de esta desde Jupyter.

## **Objetivos.**

- Creación de una base de datos.
- Creación de una vista de la base de datos.
- Creación de un procedimiento almacenado de la base de datos.
- Creación de una función de la base de datos.
- Conexión a la base de datos desde Jupyter.
- Análisis de la consulta de datos.

## Descripción de fuente de datos.

En este proyecto se hará uso de la información sobre crímenes cometidos en México, por entidad, de 2015 a agosto de 2022 disponibles en los Datos abiertos de Incidencia delictiva (SESNSP) (2022).

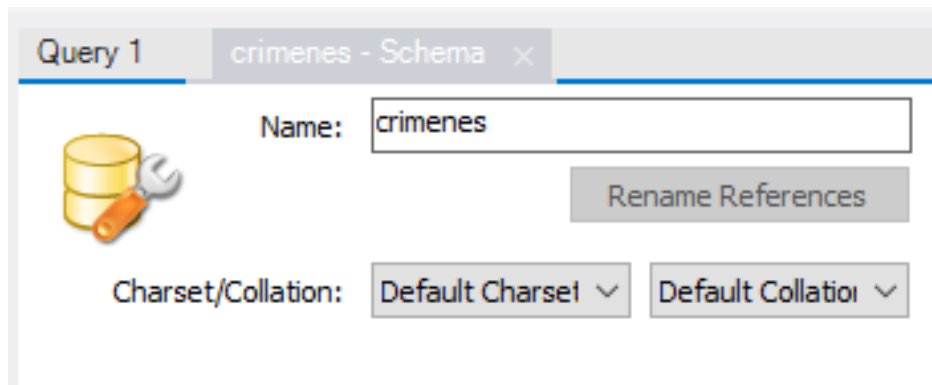
Los datos utilizados se describen en la Tabla 1.

Variable	Etiqueta	Descripción de la variable	Tipo de variable
Año	Año	Año de registro de las averiguaciones previas y/o carpetas de investigación.	Int
Clave_Ent	Clave de la entidad	Clave de la entidad, según el Marco Geoestadístico Nacional (MGN) del Instituto Nacional de Geografía y Estadística (INEGI).	Int
Entidad	Entidad	Entidad federativa de registro de las averiguaciones previas y/o carpetas de investigación.	String
Bien jurídico afectado	Bien jurídico afectado	Primera clasificación de los delitos en las averiguaciones previas y/o carpetas de investigación.	String
Tipo de delito	Tipo de delito	Segunda clasificación de los delitos.	String
Subtipo de delito	Subtipo de delito	Tercera clasificación de los delitos.	String
Modalidad	Cómo se realizó el delito	Cuarta clasificación de los delitos.	String
Total	Total	Total de registros de las averiguaciones previas y/o carpetas de investigación.	Int

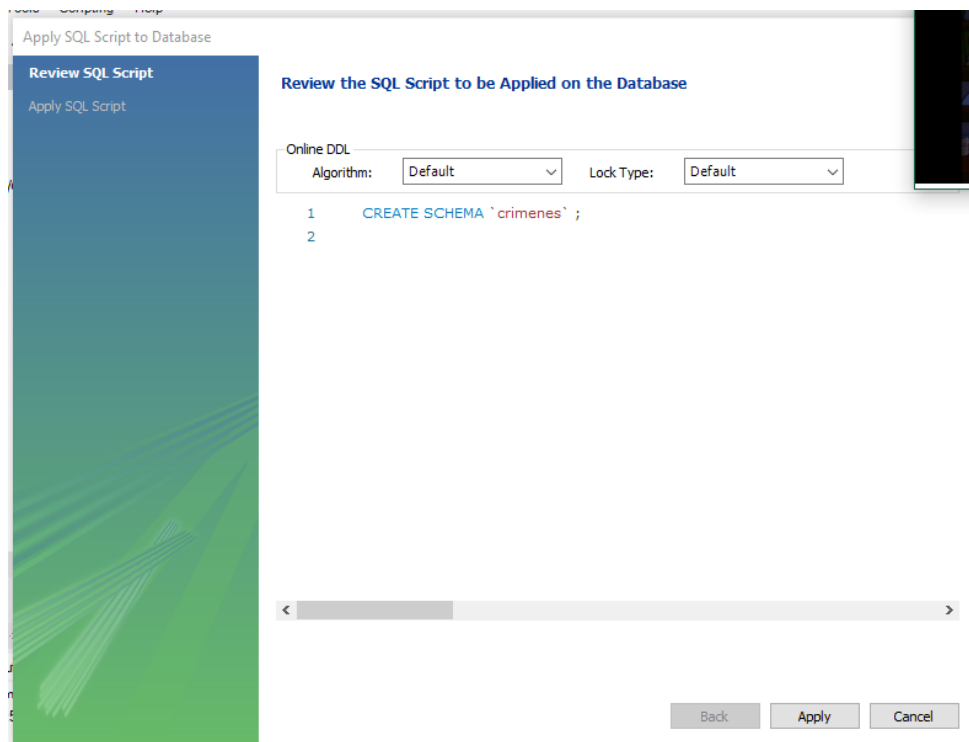
# 1. Manejo de datos en MySQL Workbench.

## 1.1 Creación de base de datos.

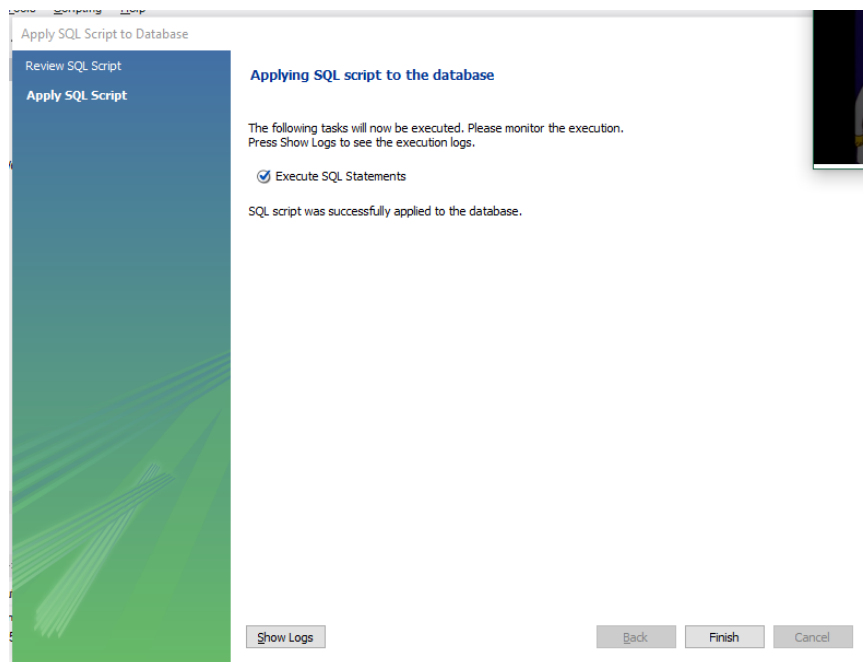
En MySQL Workbench se crea una base de datos con el nombre 'crimenes'.



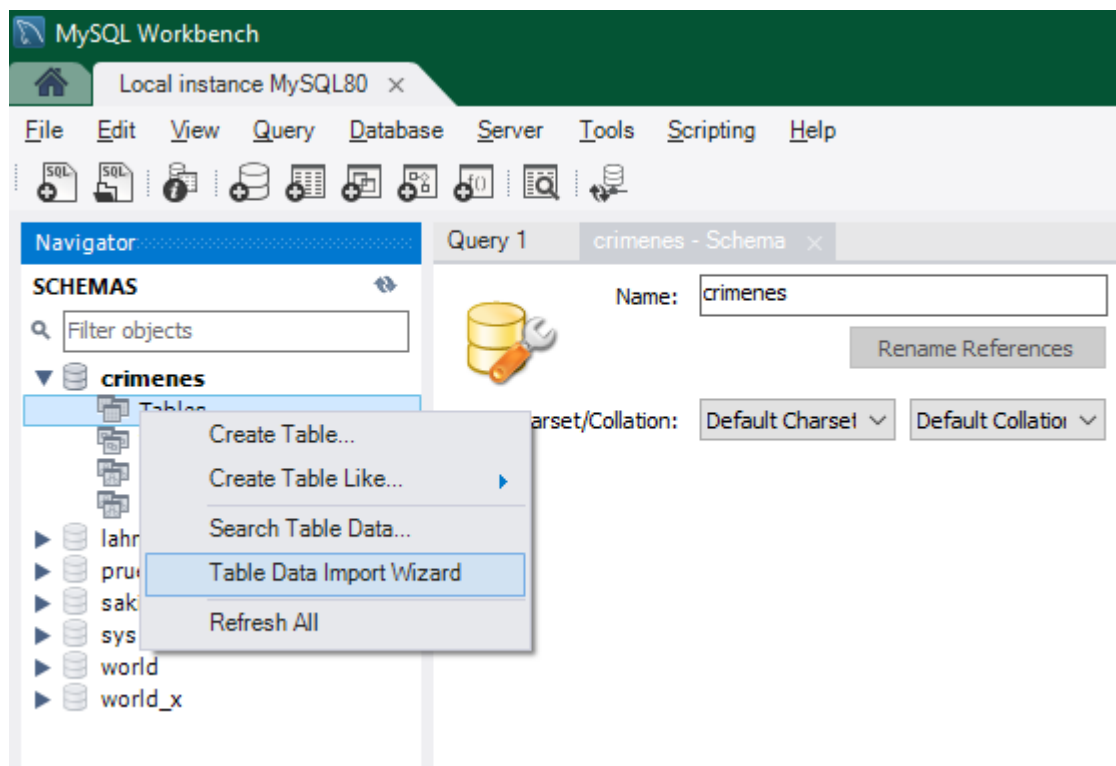
Así es el proceso de la creación de la base de datos.



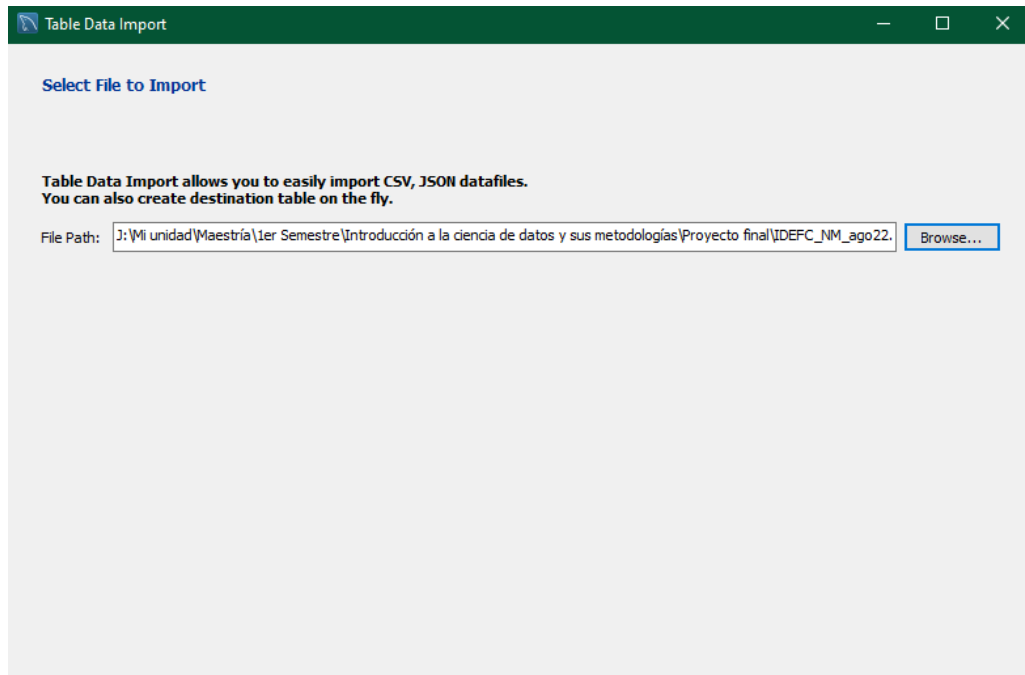
Aparece esta ventana, la cual indica que el proceso ha terminado.



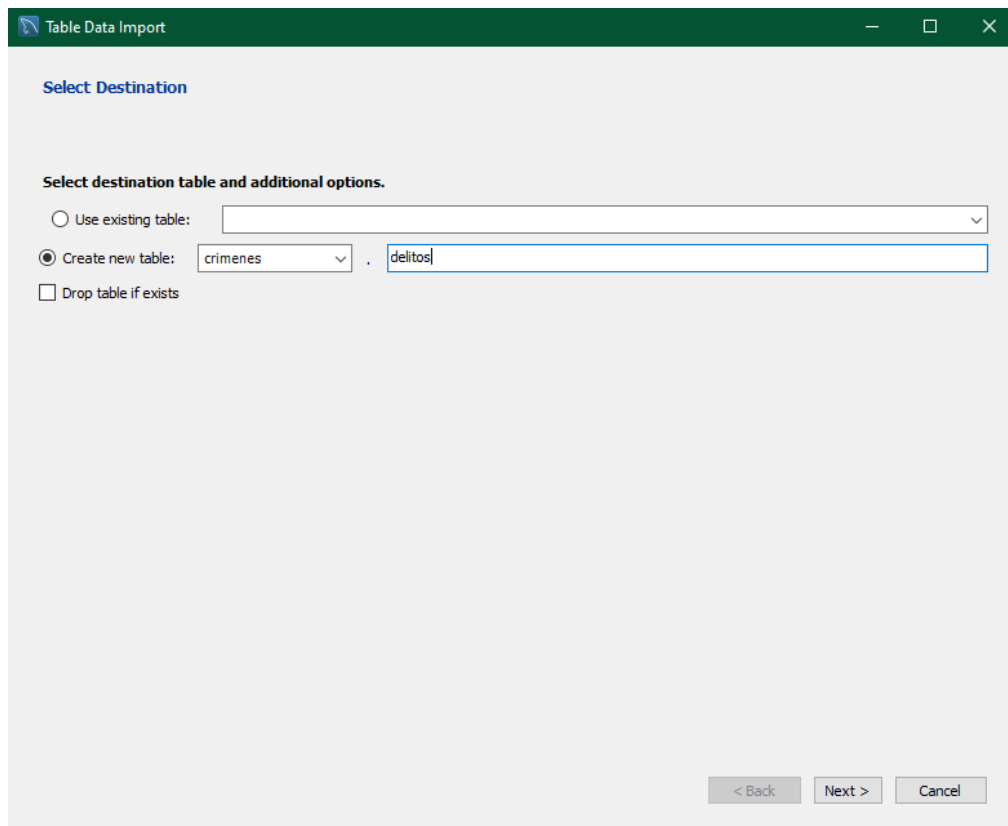
Para poder crear una tabla, se utilizó el wizard de MySQL.



Se seleccionó el archivo CSV.



Una vez seleccionado el archivo con el que se trabajó, se nombró la nueva tabla con el nombre 'delitos'.



Se verificó que el tipo de cada columna fuera correcto.

The screenshot shows the 'Table Data Import' window with the 'Configure Import Settings' tab selected. The detected file format is 'csv' and the encoding is 'utf-8'. A list of columns is shown with checkboxes and field types. Below this, a preview of the data is displayed in a table format. At the bottom, there is a 'Date format' field and navigation buttons.

Columns:	Source Column	Field Type
<input checked="" type="checkbox"/>	MyUnknownColumn	int
<input checked="" type="checkbox"/>	Año	datetime
<input checked="" type="checkbox"/>	Clave_Ent	int
<input checked="" type="checkbox"/>	Entidad	text
<input checked="" type="checkbox"/>	Bien jurídico afectado	text
<input checked="" type="checkbox"/>	Tipo de delito	text

MyUnknown...	Año	Clave_Ent	Entidad	Bien jurídic...	Tipo de delito	Subtipo de...	Modalidad	Total
0	2015	1	Aguascalien...	La vida y la...	Homicidio	Homicidio d...	Con arma d...	18
1	2015	1	Aguascalien...	La vida y la...	Homicidio	Homicidio d...	Con arma b...	5
2	2015	1	Aguascalien...	La vida y la...	Homicidio	Homicidio d...	Con otro el...	12
3	2015	1	Aguascalien...	La vida y la...	Homicidio	Homicidio d...	No especifi...	3
4	2015	1	Aguascalien...	La vida y la...	Homicidio	Homicidio c...	Con arma d...	1

Date format: %Y

< Back   Next >   Cancel

Se finalizó la importación en un total de 157.816 segundos, siendo un total de 25,088 registros.

The screenshot shows the 'Table Data Import' window with the 'Import Results' tab selected. It displays the file path, the time taken for import, the table name, and the number of records imported.

**Import Results**

File J:\Mi unidad\Maestría\1er Semestre\Introducción a la ciencia de datos y sus metodologías\Proyecto final\IDFC\_NM\_ago22.csv was imported in 157.816 s

Table crimenes.delitos was created

25088 records imported



Al descargar el archivo por medio de Wizard, no se colocó de forma correcta lo que es el Primary Key y la columna de index se colocó como "MyUnkownColumn".

MySQL Workbench interface showing the 'delitos' table in the 'crimenes' schema. The table structure is as follows:

MyUnknownColumn	Año	Clave_Ent	Entidad	Bien jurídico afectado	Tipo de delito	Subtipo de delito	Modalidad	Total
0	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma de fuego	18
1	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma blanca	5
2	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con otro elemento	12
3	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	No especificado	3
4	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con arma de fuego	1
5	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con arma blanca	0
6	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	En accidente de tránsito	163
7	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con otro elemento	2
8	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	No especificado	7
9	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Lesiones	Lesiones dolosas	Con arma de fuego	32
10	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Lesiones	Lesiones dolosas	Con arma blanca	152

The 'Output' pane shows the execution of the following SQL statements:

```
6 22:50:02 CREATE TABLE `crimenes`.`delitos` (`MyUnknownColumn` int, `Año` datetime, `Clave_Ent` in... OK 0.000 sec
7 22:50:02 PREPARE stmt FROM INSERT INTO `crimenes`.`delitos` (`MyUnknownColumn`,`Año`,`Clav... OK 0.000 sec
8 22:53:38 DEALLOCATE PREPARE stmt OK 0.000 sec
9 22:54:58 SELECT * FROM crimenes.delitos LIMIT 0, 50000 25088 row(s) returned 0.000 sec / 0.078 sec
```

Por esta razón, fue necesario modificarlo por medio de la opción "Alter table", se colocó la columna "MyUnknownColumn" como el primary key (PK) y se le cambió el nombre a "ID".

MySQL Workbench interface showing the 'delitos' table in the 'crimenes' schema. The table structure is being modified. The 'Table Name' is 'delitos' and the 'Schema' is 'crimenes'. The 'Column Name' is 'ID' and the 'Datatype' is 'INT'. The 'Primary Key' checkbox is checked.

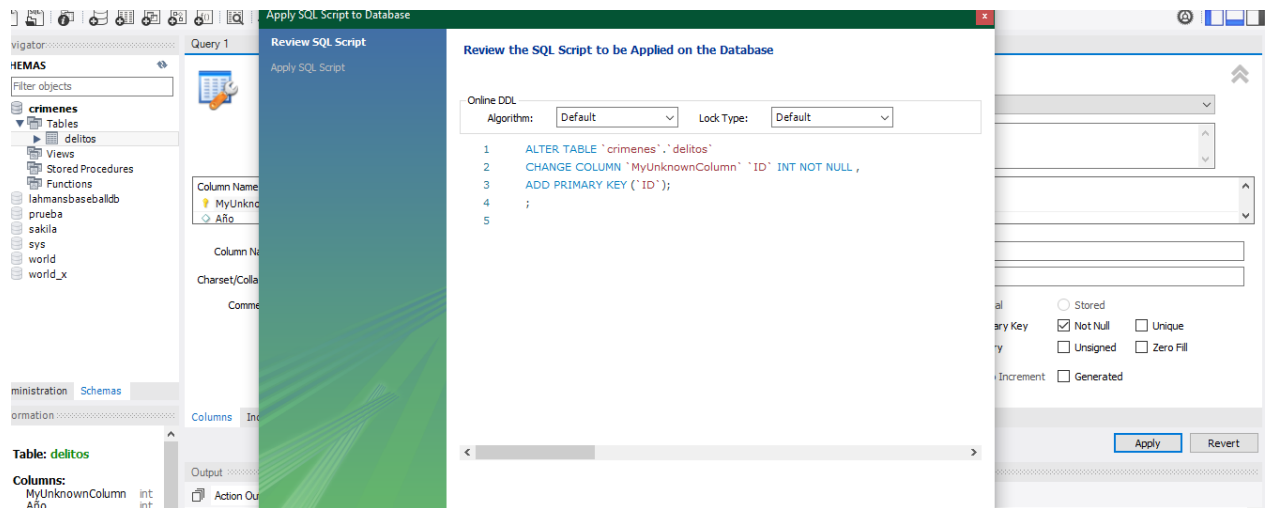
The 'Columns' pane shows the following columns:

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
MyUnknownColumn	INT	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Año	DATETIME	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

The 'Output' pane shows the execution of the following SQL statements:

```
6 22:50:02 CREATE TABLE `crimenes`.`delitos` (`MyUnknownColumn` int, `Año` datetime, `Clave_Ent` in... OK 0.000 sec
7 22:50:02 PREPARE stmt FROM INSERT INTO `crimenes`.`delitos` (`MyUnknownColumn`,`Año`,`Clav... OK 0.000 sec
8 22:53:38 DEALLOCATE PREPARE stmt OK 0.000 sec
9 22:54:58 SELECT * FROM crimenes.delitos LIMIT 0, 50000 25088 row(s) returned 0.000 sec / 0.078 sec
```

Así es como se completan los cambios.



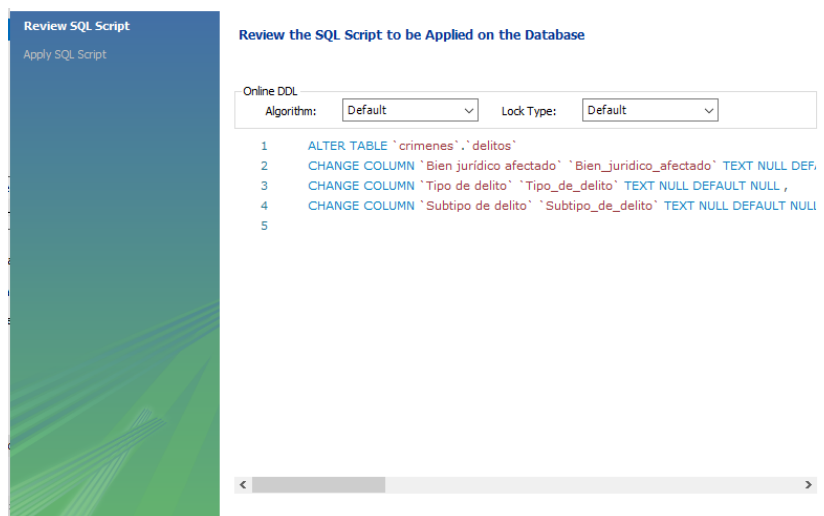
También podemos observar que las columnas “Bien jurídico afectado”, “Tipo de delito” y “Subtipo de delito” tienen espacios en el nombre. Esto provoca problemas al momento de realizar consultas. Utilizando la opción "Alter table" cambiamos los espacios por “\_”.

Column Name	Datatype	PK	NN	UQ	B
Tipo_de_delito	TEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Subtipo de delito	TEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

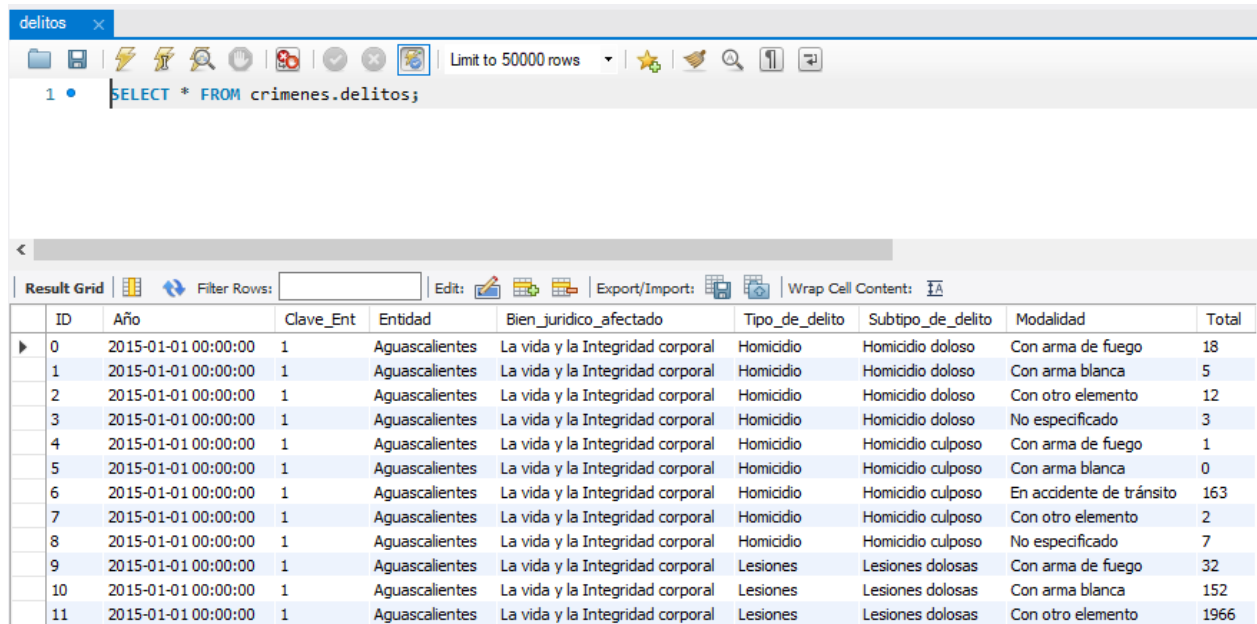
Column Name: Tipo\_de\_delito

Charset/Collation: Default Charset

Se aplican los cambios.



Así es como queda después de la actualización.

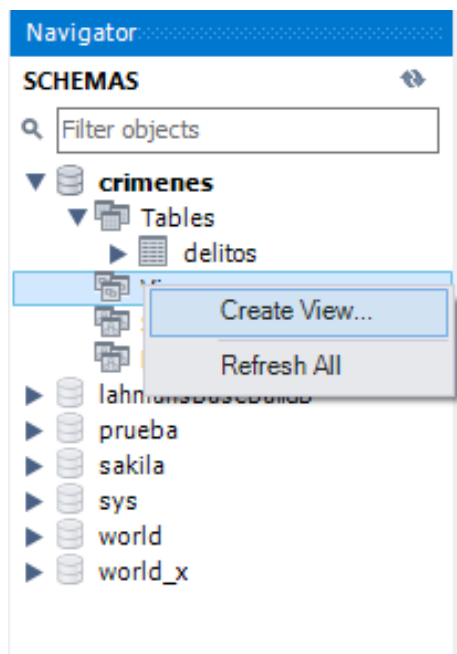


The screenshot shows a SQL client window titled 'delitos'. The query bar contains the SQL statement: `SELECT * FROM crimenes.delitos;`. Below the query bar is a 'Result Grid' with 11 rows and 10 columns. The columns are: ID, Año, Clave\_Ent, Entidad, Bien\_juridico\_afectado, Tipo\_de\_delito, Subtipo\_de\_delito, Modalidad, and Total. The data is as follows:

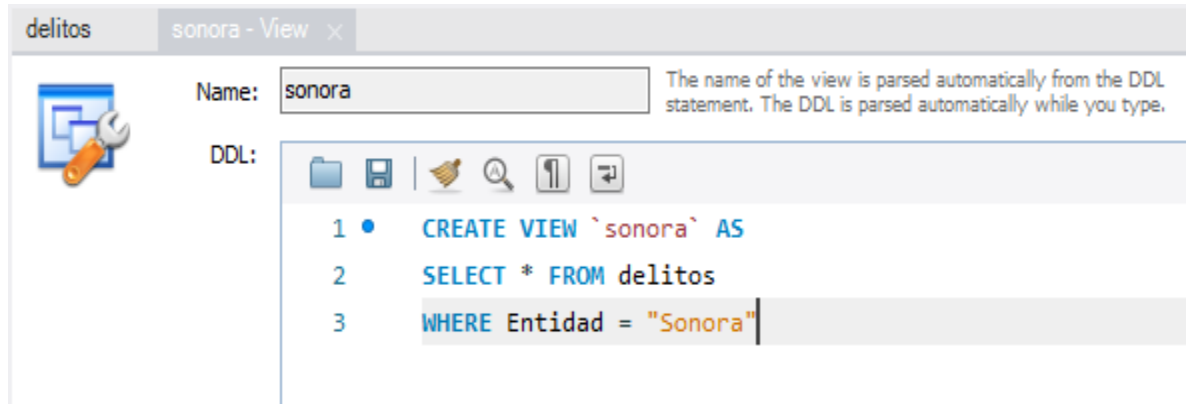
ID	Año	Clave_Ent	Entidad	Bien_juridico_afectado	Tipo_de_delito	Subtipo_de_delito	Modalidad	Total
0	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma de fuego	18
1	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma blanca	5
2	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con otro elemento	12
3	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	No especificado	3
4	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con arma de fuego	1
5	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con arma blanca	0
6	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	En accidente de tránsito	163
7	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con otro elemento	2
8	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	No especificado	7
9	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Lesiones	Lesiones dolosas	Con arma de fuego	32
10	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Lesiones	Lesiones dolosas	Con arma blanca	152
11	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Lesiones	Lesiones dolosas	Con otro elemento	1966

## 1.2 Creación de una vista.

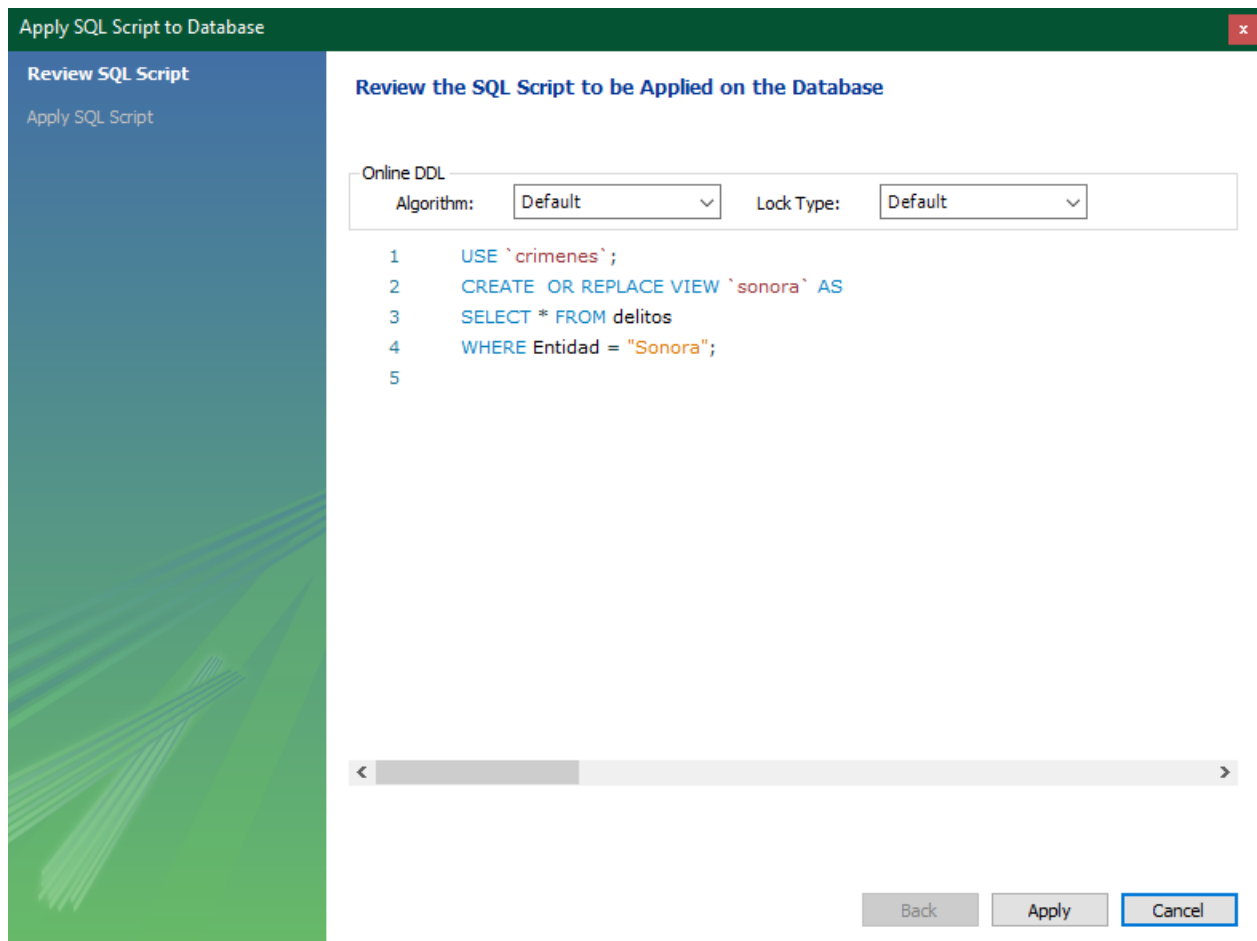
Supongamos que queremos una vista que solo nos muestre los registros de Sonora. Esto lo podemos crear realizando una vista que incluya los registros con valor de la columna Entidad igual a Sonora. Creamos la vista.



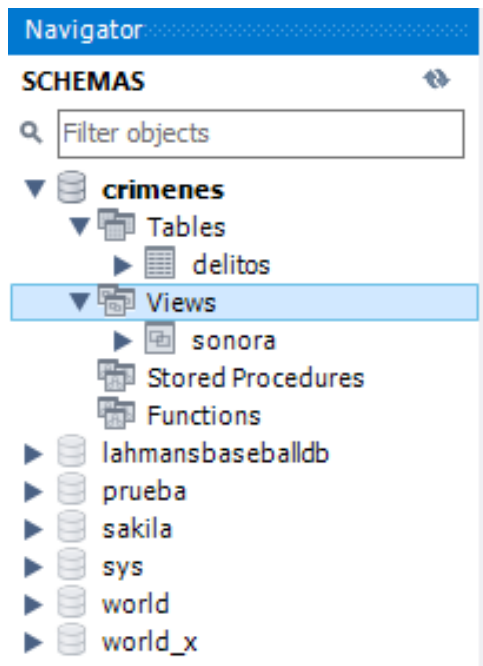
Por medio de esta consulta se obtienen los datos que interesan.



Se aplican los cambios.



La vista ya está disponible.



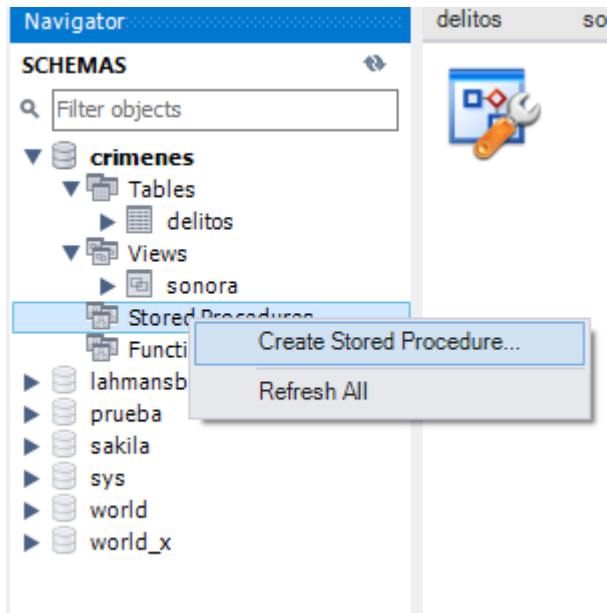
Y puede utilizarse.

The screenshot shows a SQL query window with the query `SELECT * FROM crimenes.sonora;` executed. The results are displayed in a grid with the following columns: ID, Año, Clave\_Ent, Entidad, Bien\_juridico\_afectado, Tipo\_de\_delito, Subtipo\_de\_delito, Modalidad, and Total.

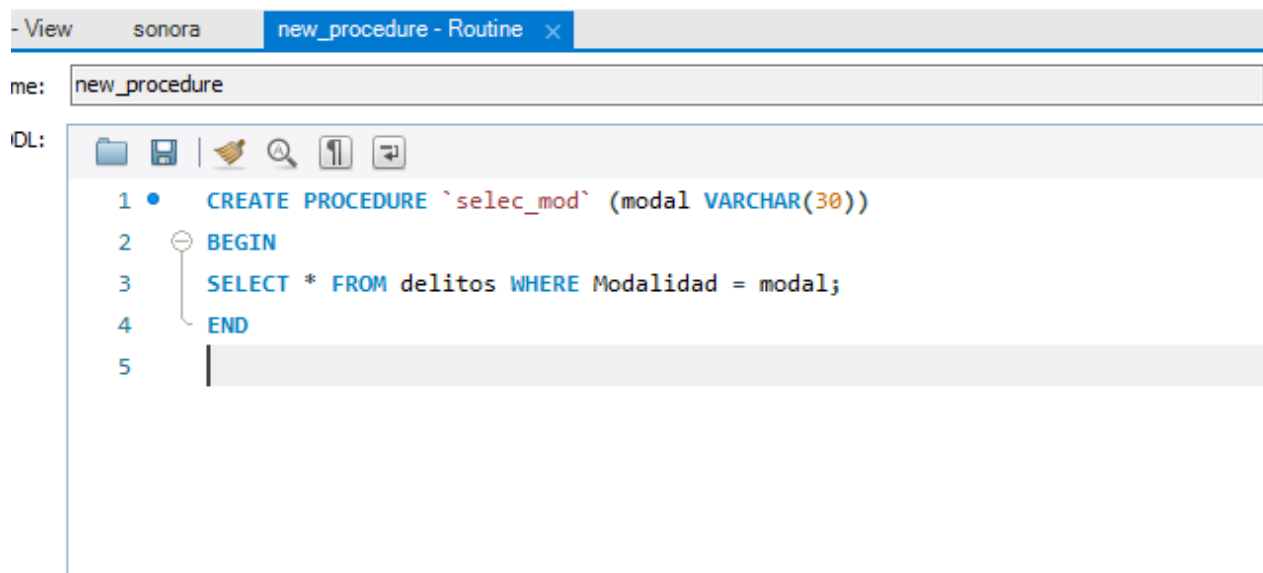
ID	Año	Clave_Ent	Entidad	Bien_juridico_afectado	Tipo_de_delito	Subtipo_de_delito	Modalidad	Total
2450	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma de fuego	341
2451	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma blanca	94
2452	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con otro elemento	154
2453	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Homicidio	Homicidio doloso	No especificado	0
2454	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con arma de fuego	0
2455	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con arma blanca	3
2456	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Homicidio	Homicidio culposo	En accidente de tránsito	292
2457	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con otro elemento	0
2458	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Homicidio	Homicidio culposo	No especificado	99
2459	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Lesiones	Lesiones dolosas	Con arma de fuego	0
2460	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Lesiones	Lesiones dolosas	Con arma blanca	0
2461	2015-01-01 00:00:00	26	Sonora	La vida y la Integridad corporal	Lesiones	Lesiones dolosas	Con otro elemento	0

### 1.3 Creación de un procedimiento almacenado.

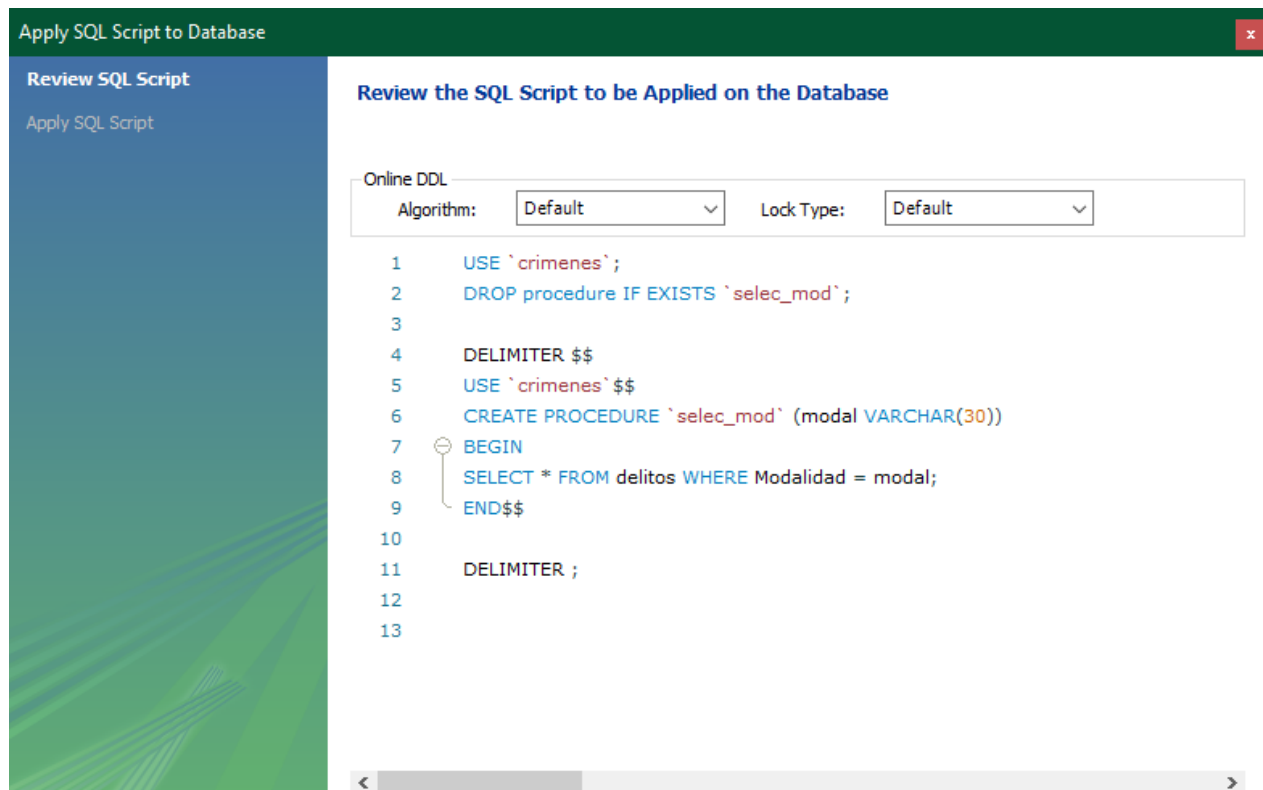
Imaginemos que de manera constante se estará realizando la consulta de delitos cometidos a partir de la modalidad. Es en este escenario donde crear un stored procedure puede ser conveniente. Procedemos a crear entonces el Stored Procedure.



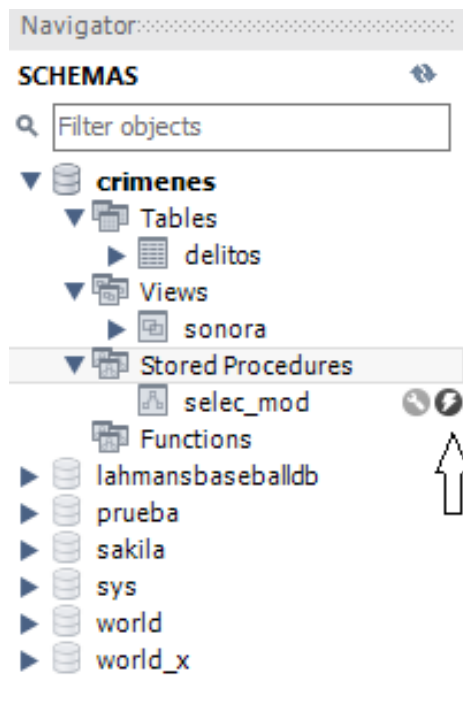
Se define el stored procedure con el siguiente código.



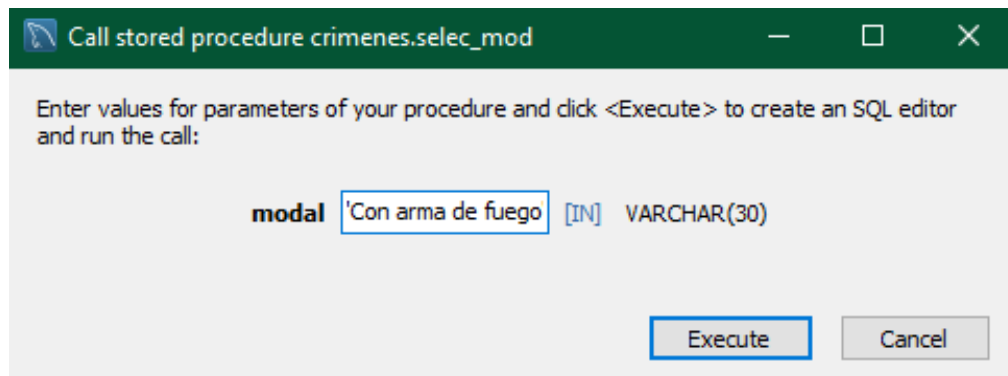
Se aplican los cambios.



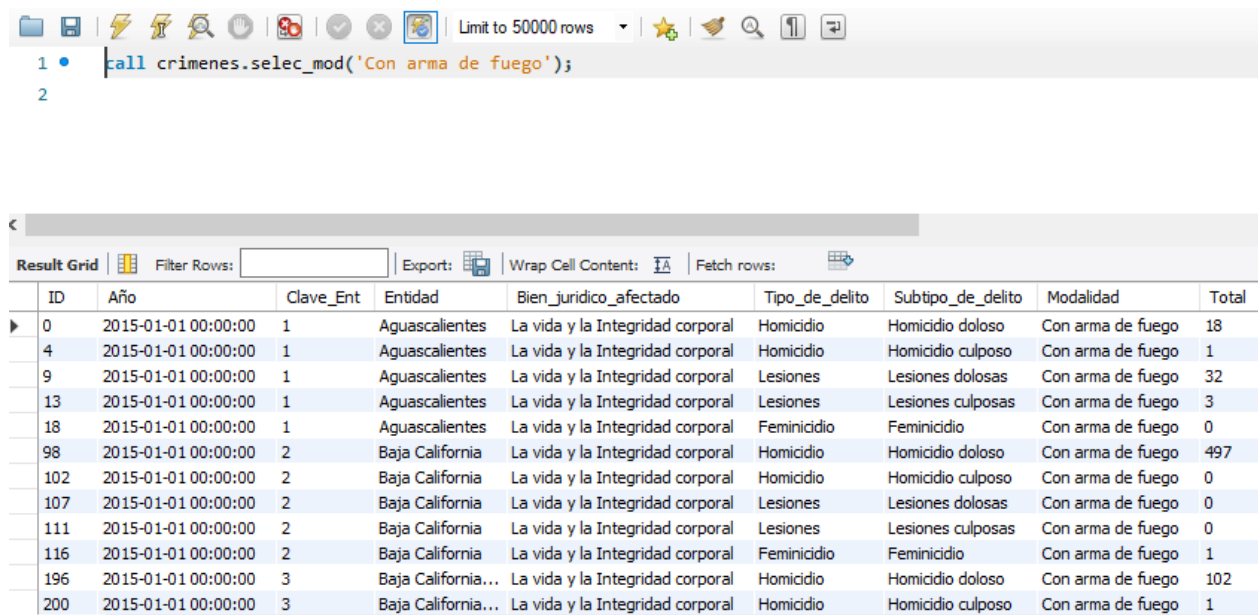
Se puede usar haciendo click en el rayito.



Se solicita que se muestren los delitos cometidos Con arma de fuego.



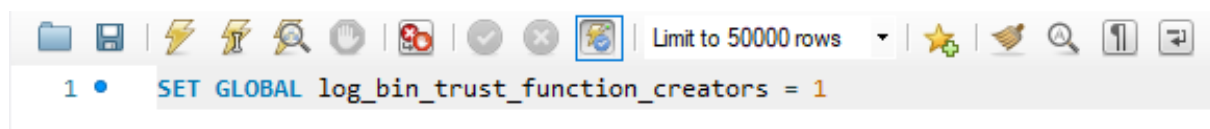
Se obtienen estos resultados.



ID	Año	Clave_Ent	Entidad	Bien_juridico_afectado	Tipo_de_delito	Subtipo_de_delito	Modalidad	Total
0	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma de fuego	18
4	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con arma de fuego	1
9	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Lesiones	Lesiones dolosas	Con arma de fuego	32
13	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Lesiones	Lesiones culposas	Con arma de fuego	3
18	2015-01-01 00:00:00	1	Aguascalientes	La vida y la Integridad corporal	Feminicidio	Feminicidio	Con arma de fuego	0
98	2015-01-01 00:00:00	2	Baja California	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma de fuego	497
102	2015-01-01 00:00:00	2	Baja California	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con arma de fuego	0
107	2015-01-01 00:00:00	2	Baja California	La vida y la Integridad corporal	Lesiones	Lesiones dolosas	Con arma de fuego	0
111	2015-01-01 00:00:00	2	Baja California	La vida y la Integridad corporal	Lesiones	Lesiones culposas	Con arma de fuego	0
116	2015-01-01 00:00:00	2	Baja California	La vida y la Integridad corporal	Feminicidio	Feminicidio	Con arma de fuego	1
196	2015-01-01 00:00:00	3	Baja California...	La vida y la Integridad corporal	Homicidio	Homicidio doloso	Con arma de fuego	102
200	2015-01-01 00:00:00	3	Baja California...	La vida y la Integridad corporal	Homicidio	Homicidio culposo	Con arma de fuego	1

#### 1.4 Creación de una función.

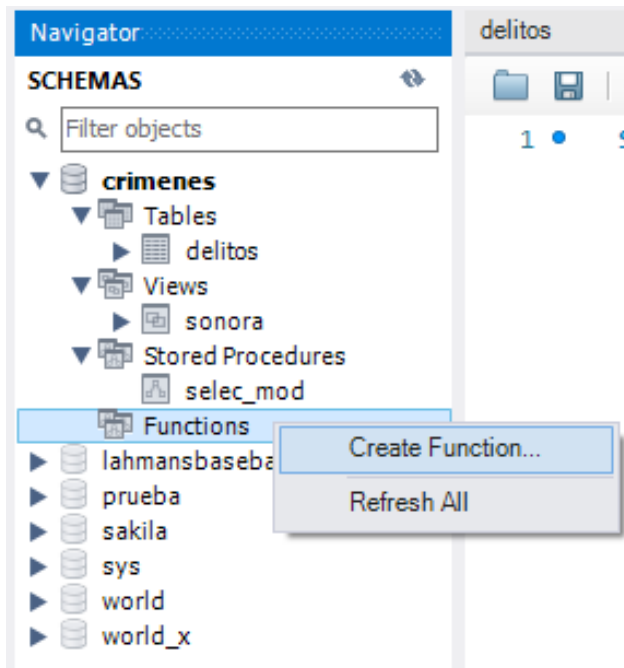
Se va a crear una función que permita introducir un tipo de delito por medio de una cadena de caracteres y regrese la cantidad de denuncias realizadas por ese tipo de delito. Antes de comenzar, es necesario correr la siguiente línea de código en la consola de SQL.



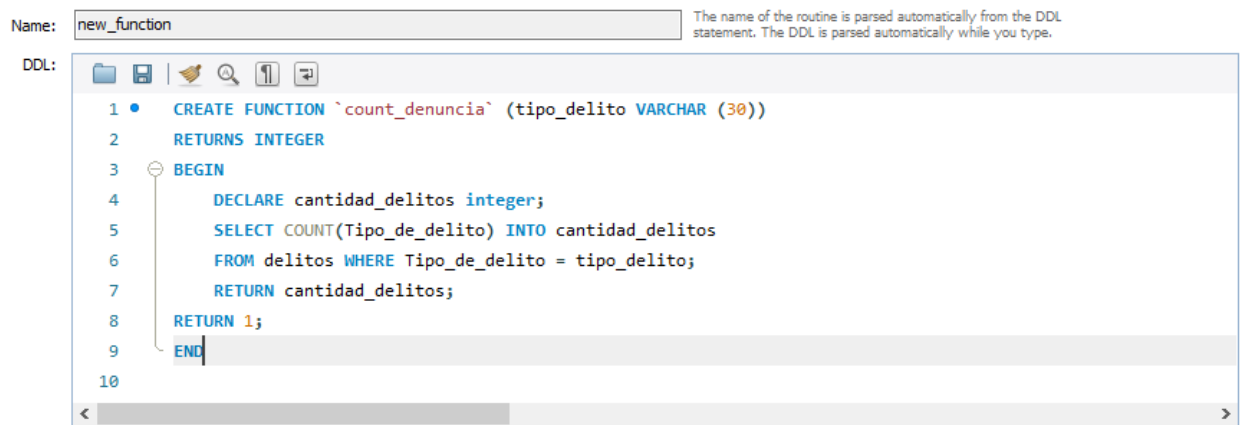
```
1 • SET GLOBAL log_bin_trust_function_creators = 1
```



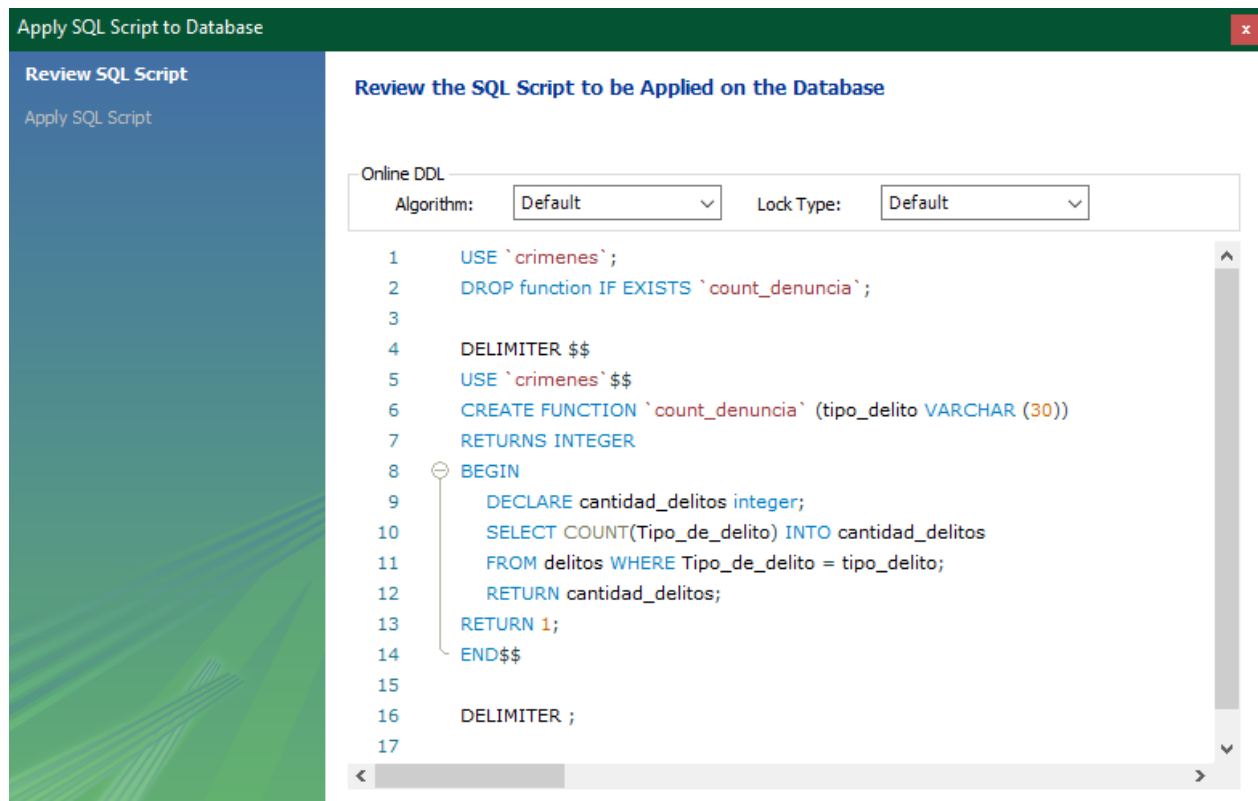
Crear función.



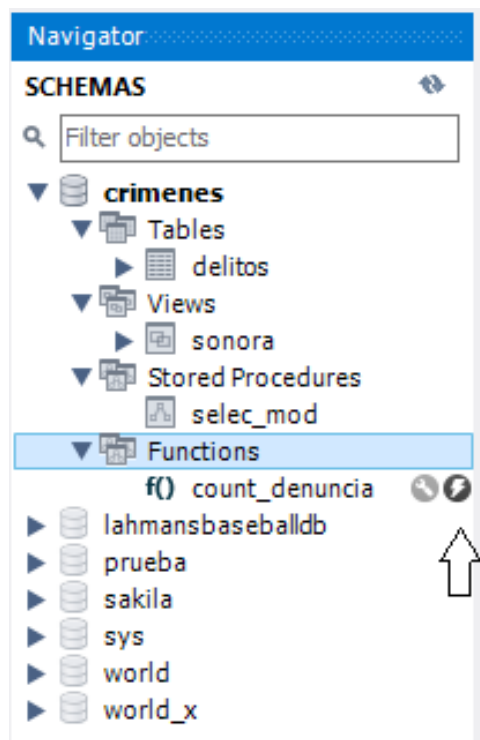
Se define la función.



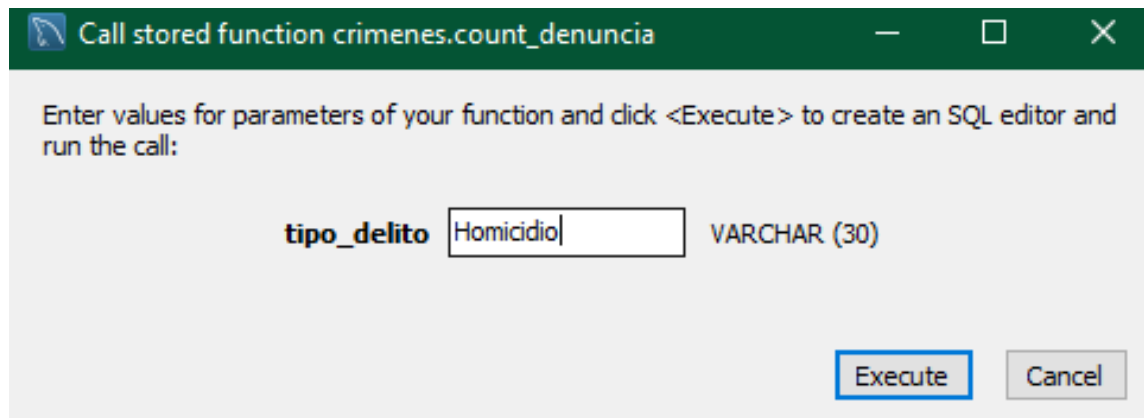
Se aplican los cambios.



Se puede usar la función haciendo click en el rayito.



Se puede pedir a la función que muestre la cantidad de Homicidios cometidos.

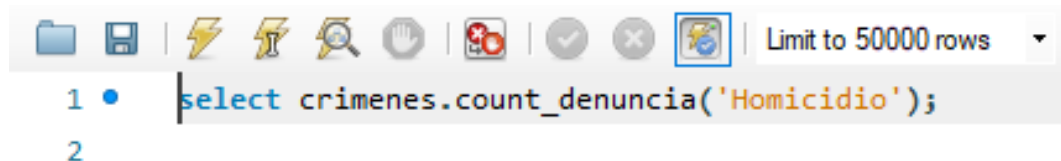


Call stored function `crimenes.count_denuncia`

Enter values for parameters of your function and click <Execute> to create an SQL editor and run the call:

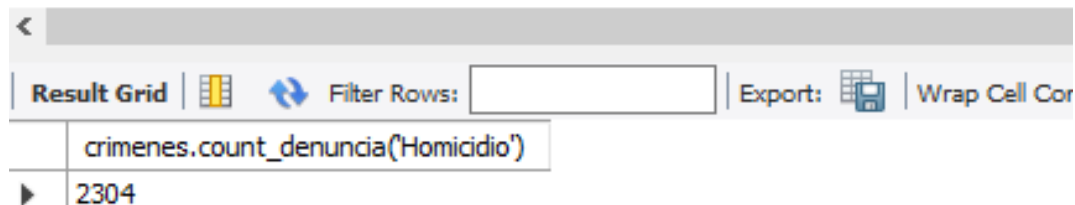
**tipo\_delito**  VARCHAR (30)

Esta es la información que despliega.



Limit to 50000 rows

```
1 • select crimenes.count_denuncia('Homicidio');
```



	crimenes.count_denuncia('Homicidio')
▶	2304

## 2. Conexión a la base de datos desde Jupyter Notebook.

Primero se importan las librerías necesarias.

```
[1]: import pymysql
import matplotlib.pyplot as plt
import pandas as pd
```

Creamos la conexión a la base de datos.

```
[2]: conn=pymysql.connect(host='localhost',
                           port=int(3306),
                           user='root',
                           password='xxx',
                           db='crimenes')
```

Realizamos una consulta para corroborar que se realizó la conexión de manera correcta.

```
[3]: dfr=pd.read_sql_query("SELECT * FROM crimenes.delitos", conn, index_col='ID')
dfr.head(5)
```

Ahora se hará una consulta más específica tomando en cuenta solo los registros referentes a Sonora.

```
[4]: query=''SELECT * FROM crimenes.delitos
      WHERE delitos.Entidad = "Sonora"'''

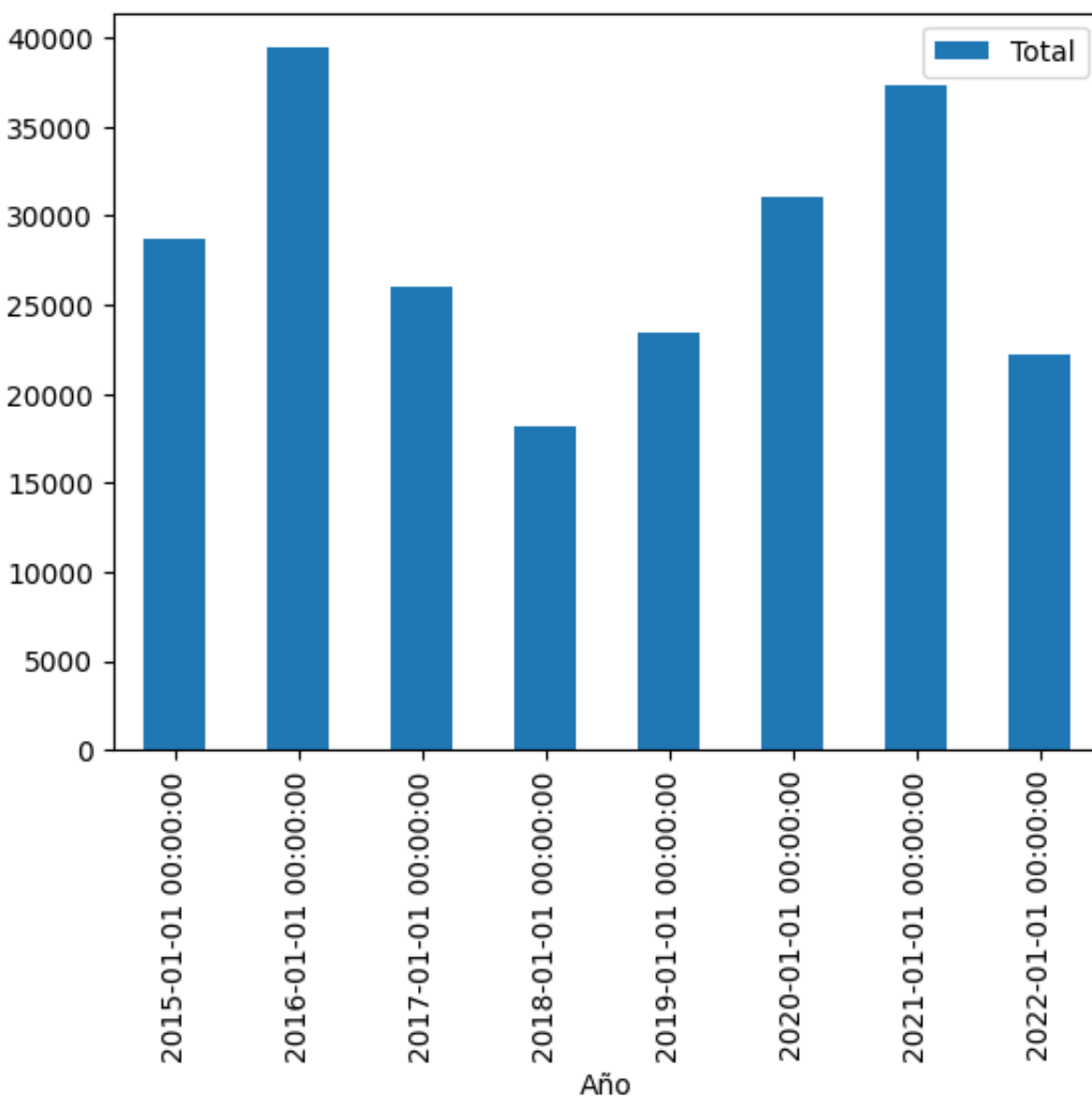
[5]: dfson=pd.read_sql_query(query, conn, index_col='ID')
dfson.head(5)
```

### 3. Análisis de la consulta de datos.

Se hizo una gráfica de barras de los años con el total de delitos denunciados en Sonora.

```
[6]: delit= dfson.groupby("Año").sum()["Total"].reset_index()  
  
delit
```

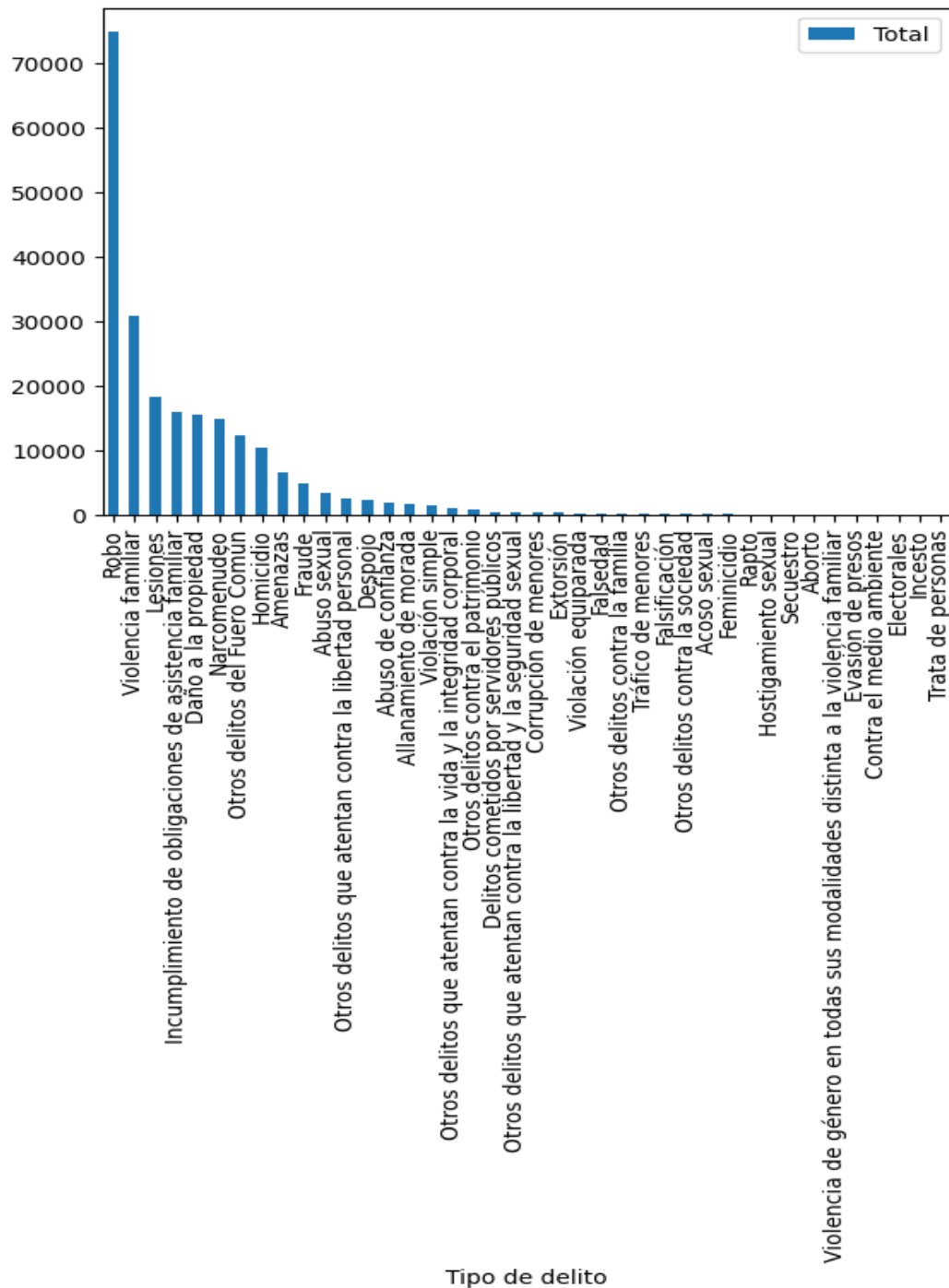
```
[7]: delit.plot.bar(x="Año",y="Total")
```



También se hizo una gráfica de barras con el total de delitos cometidos por tipo de delito en Sonora.

```
[8]: typed= dfson.groupby("Tipo de delito").sum()["Total"].reset_index().sort_values(by="Total", ascending=False)
      typed
```

```
[9]: typed.plot.bar(x="Tipo de delito",y="Total")
      plt.figure(figsize=(10, 10))
```



## 4. Repositorio de GitHub.

El proceso realizado a través de Jupyter, así como también el archivo que contiene la base de datos creada y utilizada en el proyecto, están disponibles en un repositorio de GitHub disponible [aquí](#).

## Referencias.

Gobierno de México. (15 de Septiembre de 2022). *Gobierno de México*. Obtenido de Datos Abiertos de Incidencia Delictiva: <https://www.gob.mx/sesnsp/acciones-y-programas/datos-abiertos-de-incidencia-delictiva>