

# Proyecto para examen. Sentencias SQL para analítica de datos

Objetivo. Poner en práctica las operaciones de SQL para realizar analítica de datos sobre la base de datos de incidentes viales desarrollada en prácticas previas (**VALOR INICIAL 5 PUNTOS**).

Procedimiento:

1. Retome el resultado que obtuvo en la práctica del DER de la base de datos de incidentes viales.
2. **Identifique las tablas de consulta y de los catálogos definidos (de 6 a 8).**
3. Traduzca su diagrama entidad relación. Aplique los criterios de la “Clase 31. Unidad 4. Criterios de traducción de DER a MR”.
4. Agregue del censo del 2020 de INEGI, **las características de la población total por delegación en la CDMX.** incluya información sobre **“población total femenina” y “población total masculina”** de cada delegación de la CDMX, elija algunos **rangos de la población mayor de edad.**
5. Agregue del censo del 2020 de INEGI, elija una sola opción:
  - a. Agregue el censo del 2020 por delegación: OCUPANTES POR VIVIENDA
  - b. Agregue el censo del 2020 por delegación: OCUPANTES POR CUARTO
  - c. Agregue el censo del 2020 por delegación: OCUPANTES POR DORMITORIO
6. (VALOR 2 PUNTOS) Reporte el resultado con título “. Modelo relacional de la BD de Incidentes Viales”.
7. (VALOR 2 PUNTOS) Realice la migración de datos del modelo en csv al modelo relacional, usando los incidentes viales del 2022 al 2023, también debe cargar los datos del censo del 2022 del INEGI. Debe considerar la implementación real del modelo relacional que permita la ejecución de operaciones insert, delete, y update.
8. (VALOR 1 PUNTOS) Pruebe el modelo respondiendo a las siguientes preguntas. Por cada una de ellas, documente las sentencias SQL usadas (con una breve explicación) y el resultado.
  - a. ¿Cuales son las delegaciones con el mayor número de incidentes viales (use la col. delegación de inicio)?. Grafique el resultado.
  - b.** ¿Cual es “código de cierre” con mayor frecuencia por delegación (use columna “codigo\_cierre”)?. Grafique el resultado , es decir el conteo de todos los meses.
  - c. ¿Cuales son las delegaciones con el mayor numero de población masculina y que tenga también el mayor número de incidentes viales (use la col. delegación de inicio)?. Grafique el resultado.
9. Utilice TABLEAU para graficar el resultado de las preguntas que lo ameriten.

10. (valor 1 punto) Aplique la 1era forma normal. Después aplique la 2da y/o 3era forma normal si aplica para la solución desarrollada. Reporte el resultado con título “Diagrama 3. Modelo relacional normalizado”.

**NOTA: COMO NARRATIVA Y CONOCIMIENTO DEL NEGOCIO, DEBE REVISAR EL DICCIONARIO DE DATOS DE INCIDENTES VIALES**

11. Desarrollé el documento con el diseño preliminar bajo el siguiente formato.

- Autores
- Introducción (explicación de la propuesta de solución)
- Desarrollo (incluir tablas y diagramas)
- Conclusiones (explique las razones por que las que su solución si satisface el problema)
- ENVÍE EL BACKUP DE LA BASE DE DATOS USADA Y LOS ARCHIVOS FUENTES DE TABLEAU, E IMAGENES JPG DE LOS RESULTADOS

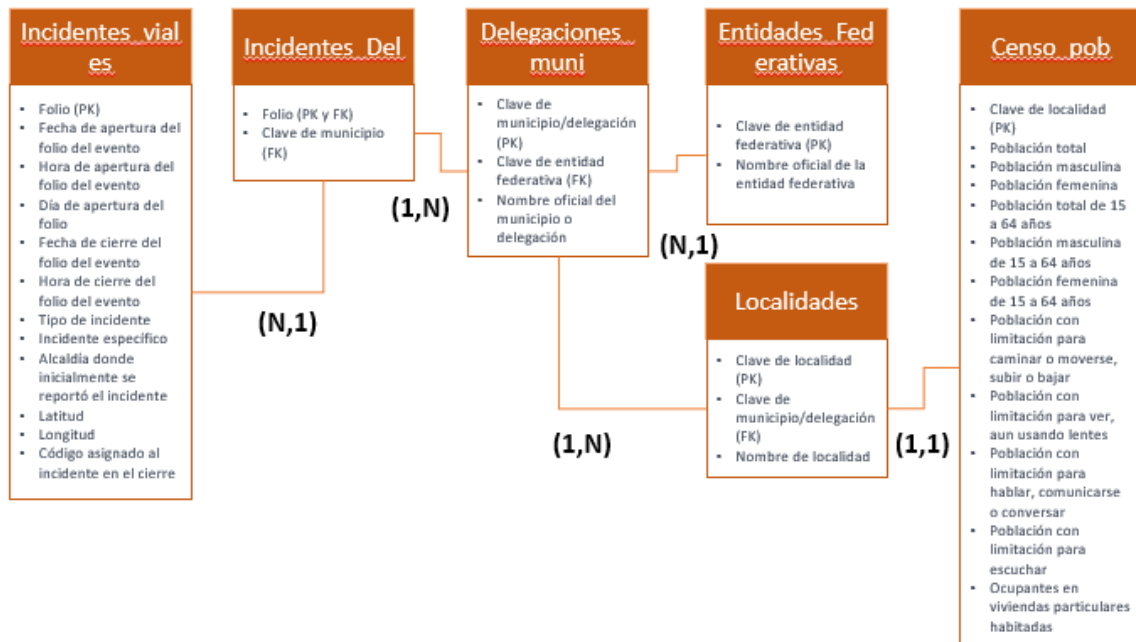
**Autores:**

**Flores Lara Alberto**

**Saucillo González Jesse Obed**

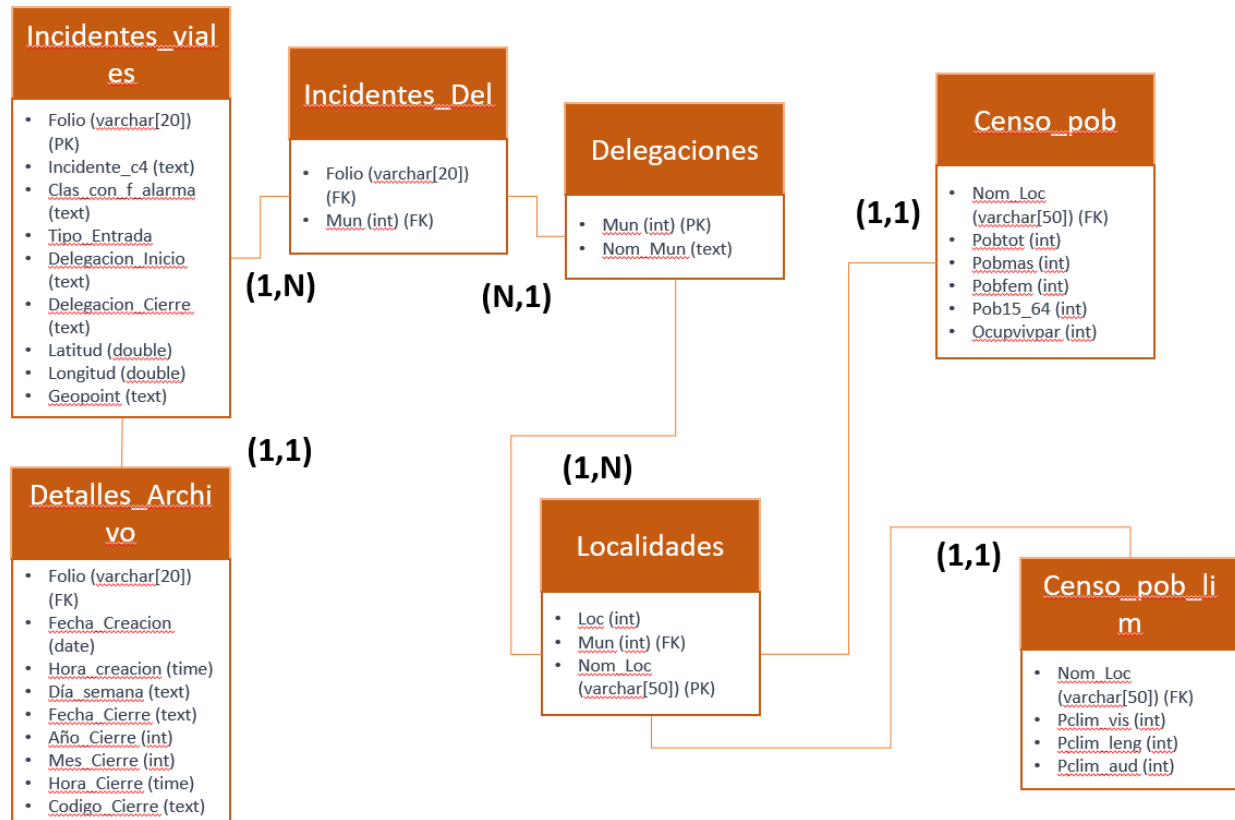
### DEL PUNTO 1-7:

Para comprender y analizar estos conjuntos de datos de manera efectiva, es fundamental desglosarlos en categorías muy específicas: El incidente vial, a que delegación pertenece el incidente, las localidades que pertenecen a las delegaciones, y las características de la población en cada localidad, basándonos en los requisitos del cliente.



Este fue el **primer boceto** que realizamos para dar solución al problema, sin embargo, nos dimos cuenta de que “Entidades\_Federativas” realmente no aporta nada a la solución del problema, y que las tablas “Incidentes\_Viales” y “Censo\_Poblacion” se encontraban muy saturadas; la estructura lógica no era la correcta y había que eliminar muchas redundancias. Por lo que **comenzamos con el proceso de normalización**, aplicando la primera, segunda y tercera forma normal, para obtener nuestro modelo relacional definitivo:

## PUNTO 9:



## DESCRIPCION DE LAS TABLAS

- **Incidentes\_Viales:** En esta tabla los atributos describen a grandes rasgos el incidente, el lugar, el tipo de incidente y su clasificación.
  - ❖ Folio (varchar[20]) (PK)
  - ❖ Incidente\_c4 (text)
  - ❖ Clas\_con\_f\_alarma (text)
  - ❖ Tipo\_Entrada
  - ❖ Delegacion\_Inicio (text)
  - ❖ Delegacion\_Cierre (text)
  - ❖ Latitud (double)

- ❖ Longitud (double)
  - ❖ Geopoint (text)
- **Detalles\_Archivo:** Se describe de forma mas extensa y explicita el tiempo en el que fueron redactados los folios de cada uno de los incidentes.
- ❖ Folio (varchar[20]) (FK)
  - ❖ Fecha\_Creacion (date)
  - ❖ Hora\_creacion (time)
  - ❖ Día\_semana (text)
  - ❖ Fecha\_Cierre (text)
  - ❖ Año\_Cierre (int)
  - ❖ Mes\_Cierre (int)
  - ❖ Hora\_Cierre (time)
  - ❖ Codigo\_Cierre (text)
- **Incidentes\_Del:** En esta tabla conectamos los registros de cada incidente con la clave de la delegación a la que corresponde.
- ❖ Folio (varchar[20]) (FK)
  - ❖ Mun (int) (FK)
- **Delegaciones:** En esta tabla, mostramos cual es la clave a la que pertenece cada una de las delegaciones de la ciudad de México.
- ❖ Mun (int) (PK)
  - ❖ Nom\_Mun (text)
- **Localidades:** Muestra cada una de las localidades que pertenecen a cada delegación.
- ❖ Loc (int)
  - ❖ Mun (int) (FK)
  - ❖ Nom\_Loc (varchar[50]) (PK)
- **Censo\_Pob:** Describe a la plobacion de cada localidad, muestra la cantidad de personas que viven en cada localidad utilizando diferentes rangos de edad y características particulares de la población como el sexo y la cantidad de personas que viven en una vivienda.
- ❖ Nom\_Loc (varchar[50]) (FK)
  - ❖ Pobtot (int)
  - ❖ Pobmas (int)
  - ❖ Pobfem (int)
  - ❖ Pob15\_64 (int)

- ❖ Ocupvivpar (int)
- **Censo\_pob\_lim:** Consideramos importante tomar en cuenta a la población que posee alguna discapacidad que aumente el riesgo de que pueda causar un accidente vial.
  - ❖ Nom\_Loc (varchar[50]) (FK)
  - ❖ Pclim\_vis (int)
  - ❖ Pclim\_leng (int)
  - ❖ Pclim\_aud (int)

## CARDINALIDADES

- “Incidentes\_Viales” y “Detalles\_Archivo” tienen una cardinalidad 1:1 porque cada registro de incidente contiene una descripción única, y cada descripción corresponde a un único registro.
- “Incidentes\_Viales” y “Incidentes\_Del” poseen una cardinalidad 1: N porque uno o varios incidentes pueden reportarse en una delegación, y solo puede reportarse una delegación de inicio par cada incidente.
- “Incidentes\_Del” y “Delegaciones” tienen una cardinalidad N:1 porque hay una delegación de inicio para cada delegación, pero cada delegación pueden ser el inicio de varios incidentes.
- “Delegaciones” tiene una cardinalidad de 1: N con la tabla "Localidades", ya que un municipio o delegación puede tener múltiples localidades, pero una localidad pertenece a un único municipio o delegación.
- "Localidades" tiene una cardinalidad de 1: 1 con la tabla "Censo\_Pob", ya que cada registro en la tabla "Localidades" se relaciona con un único registro de población en la tabla "Censo\_Pob".
- "Localidades" tiene una cardinalidad de 1: 1 con la tabla "Censo\_Pob\_Lim", ya que cada registro en la tabla "Localidades" se relaciona con un único registro de población en la tabla "Censo\_Pob\_Lim".

Conservamos los datatype que venían por default en los datasets, haciendo algunos cambios para tener los datos mas ordenados, como pasar varias fechas y horas a date y time respectivamente o cambiar algunos valores “text” a “int”. El principal cambio se los llevaron los atributos que fueron destinados a ser Primal Key y Foreign Key, ya que la mayoría eran datos del tipo “text”, tuvieron que ser modificados a tipo “Varchar” para no tener ningún problema al vincular las tablas.

## COMANDOS UTILIZADOS EN MYSQL PARA REALIZAR LA TRADUCCIÓN DEL DIAGRAMA DER AL MODELO RELACIONAL

```
create database proyecto;
```

```
Use proyecto;
```

```
CREATE TABLE proyecto.Incidentes_Viales AS  
SELECT folio, incidente_c4, latitud, longitud, clas_con_f_alarma, tipo_entrada,  
delegacion_inicio, delegacion_cierre, geopoint  
FROM censo.incidentevial;
```

```
CREATE TABLE proyecto.Detalles_Archivo AS  
SELECT folio, fecha_creacion, hora_creacion, dia_semana, codigo_cierre, fecha_cierre,  
año_cierre, mes_cierre, hora_cierre  
FROM censo.incidentevial;
```

```
CREATE TABLE proyecto.Incidentes_Del AS  
SELECT folio, delegacion_inicio  
FROM censo.incidentevial;
```

```
CREATE TABLE proyecto.Delegaciones AS  
SELECT distinct mun, nom_mun  
FROM censo.datos;
```

```
CREATE TABLE proyecto.Localidades AS  
SELECT loc, mun, nom_loc  
FROM censo.datos;
```

```
CREATE TABLE proyecto.Censo_Pob AS  
SELECT nom_loc, pobtot, pobmas, pobfem, pob15_64, ocupvivpar  
FROM censo.datos;
```

```
CREATE TABLE proyecto.Censo_Pob_Lim AS  
SELECT nom_loc, pclim_vis, pclim_leng, pclim_aud  
FROM censo.datos;
```

### **Tabla: Incidentes\_Viales**

```
UPDATE proyecto.Incidentes_Viales  
SET delegacion_inicio = CASE  
    WHEN delegacion_inicio = 'ALVARO OBREGON' THEN 'Álvaro Obregón'  
    WHEN delegacion_inicio = 'AZCAPOTZALCO' THEN 'Azcapotzalco'  
    WHEN delegacion_inicio = 'BENITO JUAREZ' THEN 'Benito Juárez'
```

```

WHEN delegacion_inicio = 'COYOACAN' THEN 'Coyoacán'
WHEN delegacion_inicio = 'CUAJIMALPA' THEN 'Cuajimalpa de Morelos'
WHEN delegacion_inicio = 'CUAUHTEMOC' THEN 'Cuauhtémoc'
WHEN delegacion_inicio = 'GUSTAVO A. MADERO' THEN 'Gustavo A. Madero'
WHEN delegacion_inicio = 'IZTACALCO' THEN 'Iztacalco'
WHEN delegacion_inicio = 'IZTAPALAPA' THEN 'Iztapalapa'
WHEN delegacion_inicio = 'MAGDALENA CONTRERAS' THEN 'La Magdalena Contreras'
WHEN delegacion_inicio = 'MIGUEL HIDALGO' THEN 'Miguel Hidalgo'
WHEN delegacion_inicio = 'MILPA ALTA' THEN 'Milpa Alta'
WHEN delegacion_inicio = 'TLALPAN' THEN 'Tlalpan'
WHEN delegacion_inicio = 'TLAHUAC' THEN 'Tláhuac'
WHEN delegacion_inicio = 'VENUSTIANO CARRANZA' THEN 'Venustiano Carranza'
WHEN delegacion_inicio = 'XOCHIMILCO' THEN 'Xochimilco'
ELSE delegacion_inicio
END;

```

UPDATE proyecto.Incidentes\_Viales

SET delegacion\_cierre = CASE

```

WHEN delegacion_cierre = 'ALVARO OBREGON' THEN 'Álvaro Obregón'
WHEN delegacion_cierre = 'AZCAPOTZALCO' THEN 'Azcapotzalco'
WHEN delegacion_cierre = 'BENITO JUAREZ' THEN 'Benito Juárez'
WHEN delegacion_cierre = 'COYOACAN' THEN 'Coyoacán'
WHEN delegacion_cierre = 'CUAJIMALPA' THEN 'Cuajimalpa de Morelos'
WHEN delegacion_cierre = 'CUAUHTEMOC' THEN 'Cuauhtémoc'
WHEN delegacion_cierre = 'GUSTAVO A. MADERO' THEN 'Gustavo A. Madero'
WHEN delegacion_cierre = 'IZTACALCO' THEN 'Iztacalco'
WHEN delegacion_cierre = 'IZTAPALAPA' THEN 'Iztapalapa'
WHEN delegacion_cierre = 'MAGDALENA CONTRERAS' THEN 'La Magdalena Contreras'
WHEN delegacion_cierre = 'MIGUEL HIDALGO' THEN 'Miguel Hidalgo'
WHEN delegacion_cierre = 'MILPA ALTA' THEN 'Milpa Alta'
WHEN delegacion_cierre = 'TLALPAN' THEN 'Tlalpan'
WHEN delegacion_cierre = 'TLAHUAC' THEN 'Tláhuac'
WHEN delegacion_cierre = 'VENUSTIANO CARRANZA' THEN 'Venustiano Carranza'
WHEN delegacion_cierre = 'XOCHIMILCO' THEN 'Xochimilco'
ELSE delegacion_cierre

```

END;

alter table proyecto.incidentes\_viales

modify folio VARCHAR(20);

alter table proyecto.incidentes\_viales

add primary key (folio);

**Tabla: Detalles\_Archivo**

```
SELECT * FROM proyecto.detalles_archivo;
```

```
Alter table proyecto.detalles_archivo modify año_cierre int;
```

```
ALTER TABLE proyecto.detalles_archivo  
ADD COLUMN fecha_creacion_temp DATE;  
UPDATE proyecto.detalles_archivo  
SET fecha_creacion_temp = STR_TO_DATE(fecha_creacion, '%d/%m/%Y');  
ALTER TABLE proyecto.detalles_archivo  
DROP COLUMN fecha_creacion;  
ALTER TABLE proyecto.detalles_archivo  
CHANGE COLUMN fecha_creacion_temp fecha_creacion DATE;
```

```
ALTER TABLE proyecto.detalles_archivo  
ADD COLUMN hora_creacion_temp TIME;  
UPDATE proyecto.detalles_archivo  
SET hora_creacion_temp = CAST(hora_creacion AS TIME);  
ALTER TABLE proyecto.detalles_archivo  
DROP COLUMN hora_creacion;  
ALTER TABLE proyecto.detalles_archivo  
CHANGE COLUMN hora_creacion_temp hora_creacion TIME;
```

```
ALTER TABLE proyecto.detalles_archivo  
ADD COLUMN hora_cierre_temp TIME;
```

```
UPDATE proyecto.detalles_archivo  
SET hora_cierre_temp = CAST(hora_cierre AS TIME);  
ALTER TABLE proyecto.detalles_archivo  
DROP COLUMN hora_cierre;  
ALTER TABLE proyecto.detalles_archivo  
CHANGE COLUMN hora_cierre_temp hora_cierre TIME;
```

```
alter table proyecto.detalles_archivo  
modify folio VARCHAR(20);
```

```
alter table proyecto.detalles_archivo  
add constraint FK_Folio  
foreign key (folio) references proyecto.incidentes_viales(folio);
```

### **Tabla: Delegaciones**

```
ALTER TABLE proyecto.delegaciones  
ADD PRIMARY KEY (`mun`);
```



### **Tabla: Incidentes\_Del**

```
ALTER TABLE proyecto.incidentes_del  
ADD COLUMN mun INT;
```

```
UPDATE proyecto.incidentes_del  
SET mun = CASE  
    WHEN delegacion_inicio = 'Azcapotzalco' THEN 2  
    WHEN delegacion_inicio = 'Benito Juárez' THEN 14  
    WHEN delegacion_inicio = 'Coyoacán' THEN 3  
    WHEN delegacion_inicio = 'Cuajimalpa de Morelos' THEN 4  
    WHEN delegacion_inicio = 'Cuauhtémoc' THEN 15  
    WHEN delegacion_inicio = 'Gustavo A. Madero' THEN 5  
    WHEN delegacion_inicio = 'Iztacalco' THEN 6  
    WHEN delegacion_inicio = 'Iztapalapa' THEN 7  
    WHEN delegacion_inicio = 'La Magdalena Contreras' THEN 8  
    WHEN delegacion_inicio = 'Miguel Hidalgo' THEN 16  
    WHEN delegacion_inicio = 'Milpa Alta' THEN 9  
    WHEN delegacion_inicio = 'Tlalpan' THEN 12  
    WHEN delegacion_inicio = 'Tláhuac' THEN 11  
    WHEN delegacion_inicio = 'Venustiano Carranza' THEN 17  
    WHEN delegacion_inicio = 'Xochimilco' THEN 13  
    WHEN delegacion_inicio = 'Álvaro Obregón' THEN 10  
END;
```

```
ALTER TABLE proyecto.incidentes_del  
drop COLUMN delegacion_inicio;
```

```
alter table proyecto.incidentes_del  
modify folio VARCHAR(20);
```

```
alter table proyecto.incidentes_del  
add primary key (folio);
```

```
alter table proyecto.incidentes_del  
add constraint FK_Folio_Del  
foreign key (folio) references proyecto.incidentes_viales(folio);
```

```
alter table proyecto.incidentes_del  
add constraint FK_Mun_Inc_Del  
foreign key (`mun`) references proyecto.delegaciones(`mun`);
```

### **Tabla: Localidades**

```
Alter table proyecto.localidades modify loc int;  
Alter table proyecto.localidades modify mun int;
```

```
alter table proyecto.localidades  
modify nom_loc VARCHAR(50);
```

```
alter table proyecto.localidades  
add primary key (nom_loc);
```

```
alter table proyecto.localidades  
modify nom_loc VARCHAR(50);
```

```
DELETE FROM proyecto.localidades  
WHERE `nom_loc` IN (  
    SELECT `nom_loc`  
    FROM (  
        SELECT `nom_loc`  
        FROM proyecto.localidades  
        GROUP BY `nom_loc`  
        HAVING COUNT(*) > 1  
    ) AS duplicados  
);
```

```
alter table proyecto.localidades  
add primary key (nom_loc);
```

```
alter table proyecto.localidades  
add constraint FK_Mun_Loc  
foreign key (`mun`) references proyecto.delegaciones(`mun`);
```

### **Tabla: Censo\_Pob**

```
Alter table proyecto.censo_pob modify loc int;  
alter table proyecto.censo_pob  
modify nom_loc VARCHAR(50);
```

```
DELETE FROM proyecto.censo_pob  
WHERE `nom_loc` IN (  
    SELECT `nom_loc`  
    FROM (  
        SELECT `nom_loc`  
        FROM proyecto.censo_pob  
        GROUP BY `nom_loc`
```

```
        HAVING COUNT(*) > 1
    ) AS duplicados
);
```

```
ALTER TABLE proyecto.censo_pob
ADD CONSTRAINT FK_Nom_Loc_Censo_Pob
FOREIGN KEY (nom_loc) REFERENCES proyecto.localidades(nom_loc);
```

### **Tabla: Censo\_Pob\_Lim**

```
Alter table proyecto.censo_pob_lim modify loc int;
alter table proyecto.censo_pob_lim
modify nom_loc VARCHAR(50);
```

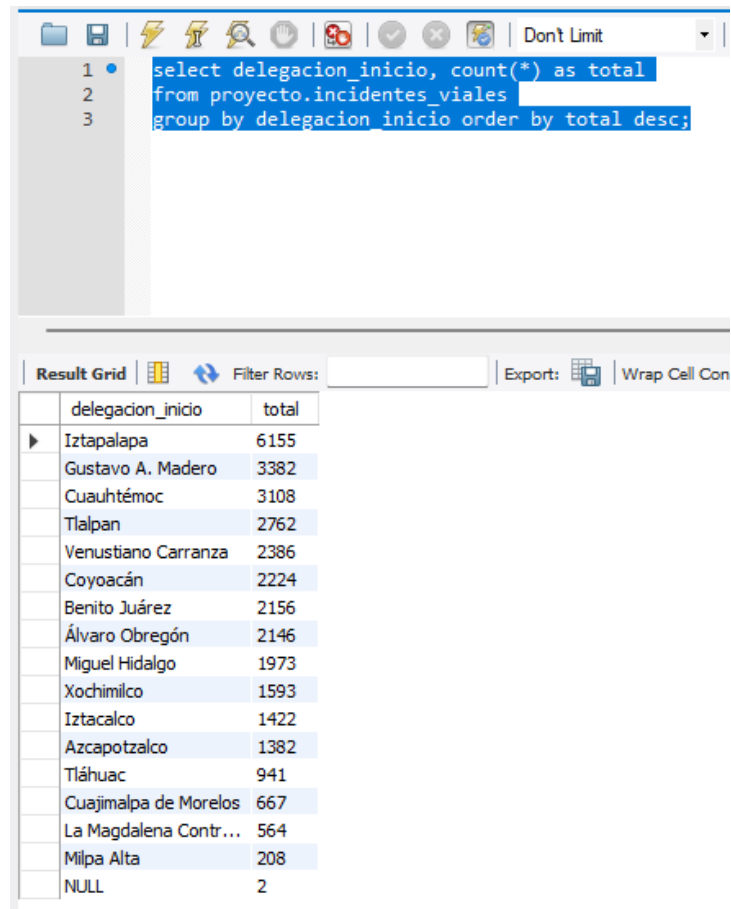
```
DELETE FROM proyecto.censo_pob_lim
WHERE `nom_loc` IN (
    SELECT `nom_loc`
    FROM (
        SELECT `nom_loc`
        FROM proyecto.censo_pob_lim
        GROUP BY `nom_loc`
        HAVING COUNT(*) > 1
    ) AS duplicados
);
```

```
ALTER TABLE proyecto.censo_pob_lim
ADD CONSTRAINT FK_Nom_Loc_Censo_Lim
FOREIGN KEY (nom_loc) REFERENCES proyecto.localidades(nom_loc);
```

## **CONSULTAS EN MYSQL**

- ¿Cuáles son las delegaciones con el mayor número de incidentes viales (use la col. delegación de inicio)?

```
select delegacion_inicio, count(*) as total
from proyecto.incidentes_viales
group by delegacion_inicio order by total desc;
```



The screenshot shows a database query tool interface. At the top, there is a toolbar with various icons for file operations, execution, and viewing. Below the toolbar, a SQL query is entered in a text area:

```

1 select delegacion_inicio, count(*) as total
2 from proyecto.incidentes_viales
3 group by delegacion_inicio order by total desc;

```

Below the query editor, the results are displayed in a table grid. The grid has two columns: 'delegacion\_inicio' and 'total'. The results are sorted in descending order of the 'total' count.

| delegacion_inicio     | total |
|-----------------------|-------|
| Iztapalapa            | 6155  |
| Gustavo A. Madero     | 3382  |
| Cuauhtémoc            | 3108  |
| Tlalpan               | 2762  |
| Venustiano Carranza   | 2386  |
| Coyoacán              | 2224  |
| Benito Juárez         | 2156  |
| Álvaro Obregón        | 2146  |
| Miguel Hidalgo        | 1973  |
| Xochimilco            | 1593  |
| Iztacalco             | 1422  |
| Azcapotzalco          | 1382  |
| Tláhuac               | 941   |
| Cuajimalpa de Morelos | 667   |
| La Magdalena Contr... | 564   |
| Milpa Alta            | 208   |
| NULL                  | 2     |

## CONCLUSIONES

Estamos completamente seguros de que nuestro modelo satisface todos los requisitos del cliente los cuales incluye datos específicos de la población y todos los datos de incidentes viales, ya que diseñamos el modelo de tal forma que no hay redundancias, cuenta con una estructura lógica donde se detalla cada tabla normalizada con sus relaciones y su cardinalidad, se pueden realizar consultas eficientes y se puede modificar la base de datos usando update o delete sin ningún problema.