Satisfying 1NF

Initial Table

DR_NO	Date Rptd	DATE OCC	TIME OCC	Area	AREA NAME	Rpt Dist No	Part	Crm Cd	Crm Cd Desc	Mocodes	Vict Age	Vict Sex	Vict Descent	Premis Cd	Premis Desc	Weapon Used Cd	Weapon Desc	Status	Status Desc	Crm Cd 1	Crm Cd 2	Crm Cd 3	Crm Cd 4	Location	Cross Street	Lat	Lon

Remove Multivalued Columns

Mocode

DR_NO	Mocodes	Description

Crime_code

DR_NO	Crm_Cd	Crm_Level	Crm Cd Desec

incident

DR_NO	Date Rptd	DATE OCC	TIME OCC	Area	AREA NAME	Rpt Dist No	Part	Vict Age	Vict Sex	Vict Descent	Premis Cd	Premis Desc	Weapon Used Cd	Weapon Desc	Status	Status Desc	Location	Cross Street	Lat	Lon

Satisfying 2NF

Mocode

DR_NO	Mocodes	Description

FD1

FD2

Crime_code

DR_NO	Crm_Cd	Crm_Level	Crm Cd Desec

FD1

FD2

Remove the Partial Dependencies

Part should go with crime_code_table

Mocode

DR_NO	Mocodes

Mocode_description

Mocodes	Description

Crime_code

DR_NO	Crm_Cd	Crm_Level

crime_code_description

Crm_Cd	Description	Part

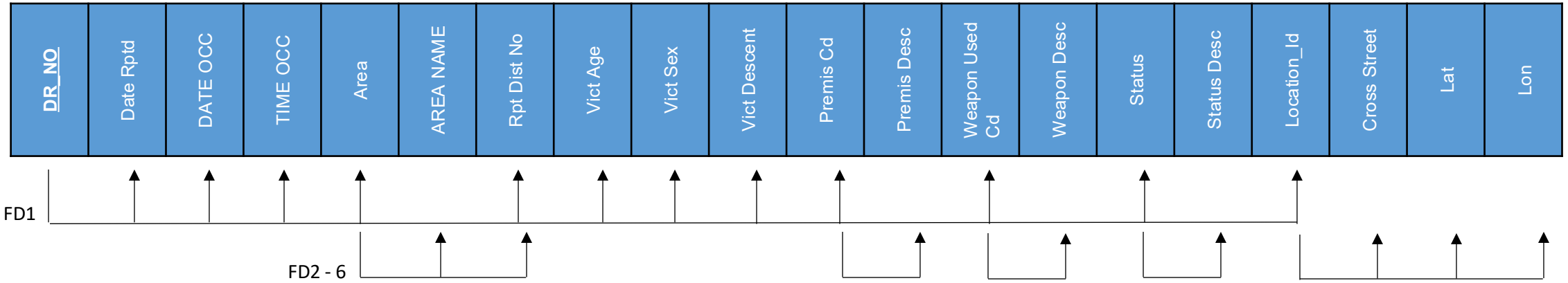
incident

DR_NO	Date Rptd	DATE OCC	TIME OCC	Area	AREA NAME	Rpt Dist No	Vict Age	Vict Sex	Vict Descent	Premis Cd	Premis Desc	Weapon Used Cd	Weapon Desc	Status	Status Desc	Location	Cross Street	Lat	Lon

FD1

Satisfying 3NF

incident



Keep Full Dependencies

incident

DR_NO	Date Rptd	Date OCC	Time OCC	Area	RPT_Dist_No	Vict Age	Vict Sex	Vict Descent	Premis Cd	Weapon Used Cd	Status	Location_Id

Decompose Transitive Dependencies

area

Area	RPT_Dist_No	Area_Name

premise

Premise Cd	Premis Desc

weapon

Weapon Used Cd	Weapon Desc

status

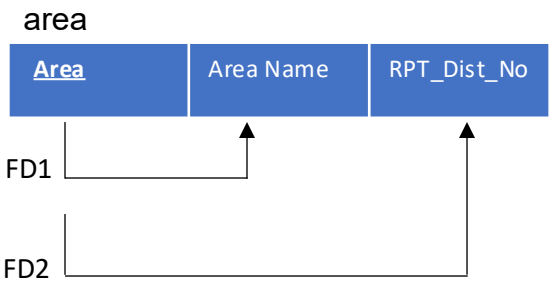
Status	Status Desc

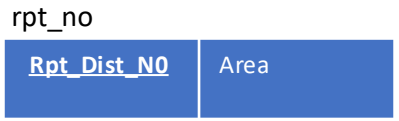
location

Location_Id	Location	Cross Street	Lat	Lon

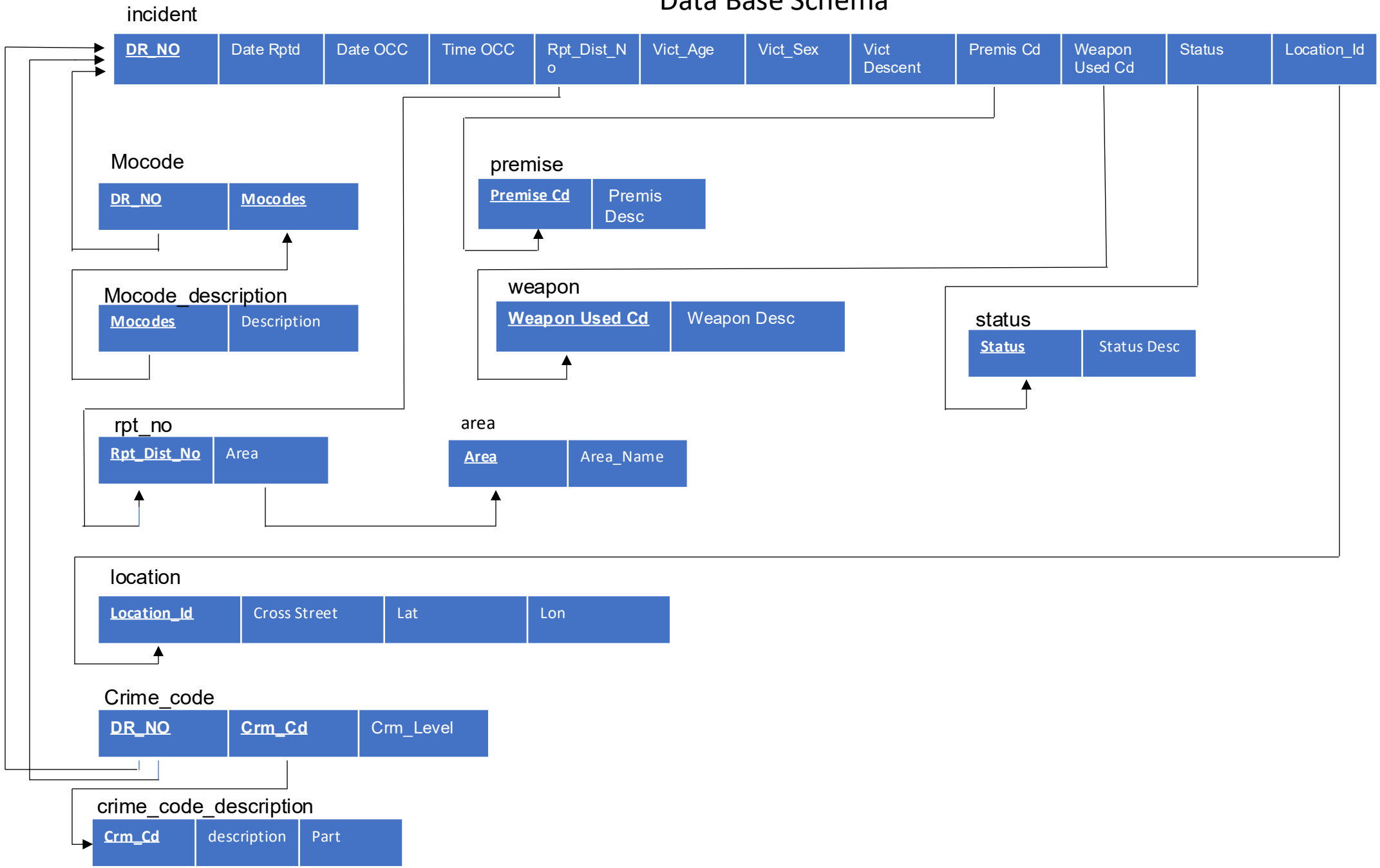
→ Surrogate key created for location for a unique identifier

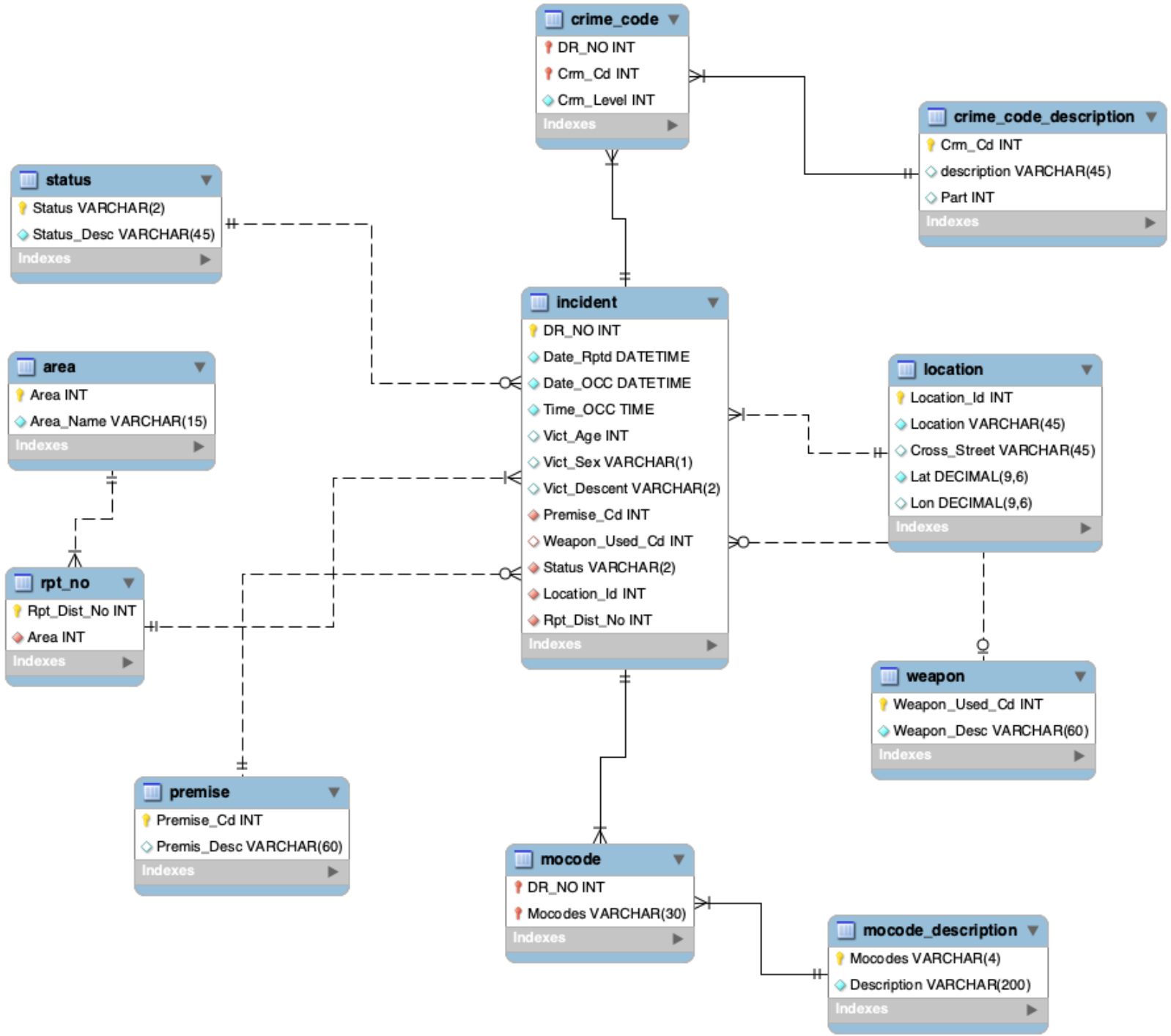
Satisfying 2NF Once More for Area Table





Data Base Schema




crime_code
DR_NO INT
Crm_Cd INT
Crm_Level INT
Indexes
crime_code_description
Crm_Cd INT
description VARCHAR(45)
Part INT
Indexes
status
Status VARCHAR(2)
Status_Desc VARCHAR(45)
Indexes
incident
DR_NO INT
Date_Rptd DATETIME
Date_OCC DATETIME
Time_OCC TIME
Vict_Age INT
Vict_Sex VARCHAR(1)
Vict_Descent VARCHAR(2)
Premise_Cd INT
Weapon_Used_Cd INT
Status VARCHAR(2)
Location_Id INT
Rpt_Dist_No INT
Indexes
area
Area INT
Area_Name VARCHAR(15)
Indexes
location
Location_Id INT
Location VARCHAR(45)
Cross_Street VARCHAR(45)
Lat DECIMAL(9,6)
Lon DECIMAL(9,6)
Indexes
rpt_no
Rpt_Dist_No INT
Area INT
Indexes
weapon
Weapon_Used_Cd INT
Weapon_Desc VARCHAR(60)
Indexes
premise
Premise_Cd INT
Premis_Desc VARCHAR(60)
Indexes
mocode
DR_NO INT
Mocodes VARCHAR(30)
Indexes
mocode_description
Mocodes VARCHAR(4)
Description VARCHAR(200)
Indexes

Assumptions Listed

1. Assuming that Rpt_Dst_No is more unique than Area or Area_Name so that is left in the incidents table.
2. Assuming that neither Location, Cross_Street, Lon, or Lat are unique enough to be a primary key.
2. A crime can happen without a weapon since some of the values for weapon are null
3. Crime code can have no description it can be null.
4. I am assuming that part should go with crime_cd_description.
5. Cross street can be null since there are missing data values.
6. There is no unique identifier for victim even if we made a surrogate key we would end up with repeated values in the database, so they are kept in the incident table.

6 Queries Planned

- **Most Frequent Crimes (Join)**
  This query identifies the top 5 most commonly reported crime types across all incidents. The tables used are incident, crime_code, and crime_code_description. I will join the incident and crime_code tables on the DR_NO column. Then, I will join the result with crime_code_description to retrieve the crime descriptions. Finally, I will count the frequency of each crime description, sort them in descending order, and display the top 5 results.

- **All Areas with or without Incidents (Outer Join)**
  This query lists all LAPD divisions along with the number of incidents reported in each, including divisions with zero incidents. I will perform a LEFT JOIN from the area table to the incident table using the AREA column. Then, I will group the results by AREA and count the number of incidents per division, ensuring all areas are displayed even if they had no incidents.

- **Mo_Codes never used in any incident (Set Theory)**
  This query identifies and displays all Modus Operandi (MO) codes that are defined in the mocode_description table but have never been used in any reported incident. This is a set difference operation: selecting all MO codes from mocode_description and subtracting those that appear in the mocode table. The result is a list of unused MO codes along with their descriptions.

- **Monthly Crime Counts per Reporting District (Aggregation with Join)**
  This query displays the number of incidents reported in each reporting district per month. The tables used are rpt_no and incident. I will join the incident table with the rpt_no table using the RPT_DIST_NO column, extract the month from DATE_OCC, and group the results by both reporting district and month.

- **Firearm related incidents**
  This query shows all incidents involving a firearm. The tables used are incident and weapon. I will join the incidentand weapon tables on the Weapon_Used_Cd column, then filter the results to include only those with firearm-related weapon descriptions such as "HANDGUN", "RIFLE", "OTHER FIREARM", and similar.

- **Firearm incidents involving female victims**
  This query shows all firearm-related incidents involving female victims, grouped by month. The tables used are incidentand weapon. I will join these tables on the Weapon_Used_Cd column and filter for records where Vict_Sex = 'F' and the weapon description includes firearm-related terms. Then, I will extract the month from DATE_OCC and group the results to count the number of such incidents per month.