

# CS434 – Data Base Theory and Design

## Project #1

### Team Database Application (TDA): Part 1 - Relational Database Design

#### Team

Lipika Baniya | 800794205 | [lbaniya@siue.edu](mailto:lbaniya@siue.edu)

The domain I would like to manage with the TDA is **Columbian Crime Datasets 2024** by the District of Columbia Metropolitan Police Department (MPD)

#### Overview

The dataset provides comprehensive information on reported crimes within a subset of locations and attributes of incidents. The purpose of the application is to create a relational database system that tracks, stores, and visualizes reported crime incidents across Washington D.C. The application will use real-world dataset provided by the District of Columbia Metropolitan Police Department (MPD). The dataset will help discover crime patterns and aid in strategic decision-making for community safety and law enforcement.

#### Entities

The “**Crimes 2024**” dataset includes the following key entities:

1. Crime\_incident
  - ccn VARCHAR PRIMARY KEY : A unique identifier assigned by MPD to each incident report.
  - report\_date TIMESTAMP : The date the offense was reported, which may be later than the date the offense actually occurred.
  - shift VARCHAR : Time the report was taken. Day shift generally runs between 0700 and 1500 (military time); evening shift between 1500 and 2300, and midnight shift between 2300 and 0700. If the shift is unknown, the field will say "UNK".
  - method\_id INT : Unique key assigned to type of weapon
  - offense\_id INT : Unique key assigned to crime offense.
  - location\_id INT : Unique key assigned to location crime occurred.
  - start\_date TIMESTAMP : Crime incident start date and time.
  - end\_date TIMESTAMP : Crime incident end date and time.
  - object\_id BIGINT : Sequential unique whole numbers that are automatically generated.
2. Method
  - method\_id INT PRIMARY KEY : Unique key assigned to type of weapon
  - Method\_name VARCHAR : Types of weapon used to commit crime.
3. Offense
  - offense\_id INT PRIMARY KEY : Unique key assigned to crime offense.
  - offense\_name VARCHAR : Crime offense.
4. Location
  - location\_id INT PRIMARY KEY : Unique key assigned to location crime occurred.

- block VARCHAR : Block Name bases on its Theoretical Address Range from MAR Geocoder.
- xblock FLOAT : Block X coordinate (centroid) of crime incident.
- yblock FLOAT : Block Y coordinate (centroid) of crime incident.
- latitude FLOAT : Latitude (decimal degrees) of Crime Incident.
- longitude FLOAT : Longitude (decimal degrees) of Crime Incident.
- geo\_id INT : Unique key assigned to geographical location.

## 5. GEOGRAPHY

- geo\_id INT PRIMARY KEY : Unique key assigned to geographical location.
- ward INT : Ward ID
- anc VARCHAR : ANC ID
- district INT : Police district
- psa INT : Police Service Areas
- neighborhood\_cluster VARCHAR : Neighborhood cluster
- block\_group VARCHAR : Census block group
- census\_tract VARCHAR : Census track
- voting\_precinct VARCHAR : Legacy Voting Precinct.
- bid VARCHAR : Business Improvement Districts.

### Relationships

Crime incident and Location have one-to-many relationship, which means one crime can have multiple location records.

Crime incident and offense have many-to-one relationship, where one offense can have multiple crimes but one crime can belong to only one offense.

Crime incident and method have many-to-one relationship, where one method can have multiple crimes but one crime can belong to only one method.

Location and Geography have one-to-many relationship, where location can belong to one geography but a geography can have multiple locations tied to it.

### Difficult Aspects

- **Missing Data** : Some rows have x\_coord = 0, y\_coord = 0.
- **Timeline Confusion** : Difference between start\_date, end\_data and report\_date
- **Geographical overlap** : A single incident may belong to overlapping jurisdictional boundaries.

### Important Queries

- Which neighbourhood clusters have the highest number of reported theft incidents?
- At what time during the day does the crime occur the most?
- Which voting precincts have experienced more than 3 assault reports at the same street block?
- Is there a certain crime types in specific wards or neighbourhoods?
- What are the top 5 blocks with the highest number of reports?
- How many incidents involved a "Gun" as the method?

# ERD

