**CS434 – Data Base Theory and Design**

**Project #4**

**Team Database Application (TDA): Part 4 – Loading Large Data Sets**

**Team**

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

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

**General Nature of application**

The main goal of an Entity Relationship Diagram (ER Diagram) is to explain the relationship between entities; it is a structural design of the database. Through the help of specialized symbols, it helps to define the relationship between entities. It is based on three main principles entities, attributes and relationships, these help to design the database that would be required before implementing the database. It is a systematic process to design a database as it would require analyzing all requirements.

**About Data**

Washington, D.C. has been facing significant challenges in ensuring public safety due to the varying and growing crime rates in different neighborhoods and time periods. It is important for law enforcement agencies to understand when and where crimes occur so that it can respond efficiently and allocate limited resources wisely. Imagine a robust database system that is designed to handle this task effectively, because without a data-driven approach and structured database, policing efforts may remain reactive, which would result in delays or gaps in coverage in high-risk areas. This database includes various entities, each representing a key component of crime data management.

**ER Diagram**

**A diagram of a crime investigation

AI-generated content may be incorrect.**

1. **Dataset**

The dataset is available in one csv file Crime\_Incidents\_in\_2024.csv

* Size: (7,306,043 bytes)
* Columns: A computer screen shot of a black screen

  AI-generated content may be incorrect.

1. **Data Cleaning and Separating into Different csv files for Each Table**

I used Python to separate the tables from the csv files. Python was used to:

* Separate the csv based on table
* Assign primary and foreign keys
* Remove duplicate values that may occur in keys

Following is a snippet of code use for data transformation:

A screen shot of a computer code

AI-generated content may be incorrect.

A screen shot of a computer code

AI-generated content may be incorrect.

A computer screen shot of text

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A computer screen with text

AI-generated content may be incorrect.

A screen shot of a computer code

AI-generated content may be incorrect.

Modified csv file on tables looked as follows:

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

1. **Adding Data into Database**

I imported data from the PostgreSQL GUI pgAdmin 4 to import csv files in bulk.

* 1. **Table Offense**

Number of tuples added: **9**

A screenshot of a computer

AI-generated content may be incorrect.

**Screenshot of Table Offense**

A screenshot of a computer

AI-generated content may be incorrect.

* 1. **Table Method**

Number of tuples: **3**

A screenshot of a computer

AI-generated content may be incorrect.

**Screenshot of Table Method**

A screenshot of a computer

AI-generated content may be incorrect.

* 1. **Table Location**

Number of tuples: **475**

A screenshot of a computer

AI-generated content may be incorrect.

**Screenshot of Table Location**

A screenshot of a computer

AI-generated content may be incorrect.

* 1. **Table Block\_Group**

Number of tuples: **8102**

**A screenshot of a computer error

AI-generated content may be incorrect.**

**Screenshot of Block\_Group**

**A screenshot of a computer

AI-generated content may be incorrect.**

* 1. **Table Crime\_Location**

Number of tuples: **29281**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Screenshot of Table Crime\_Location**

**A screenshot of a computer

AI-generated content may be incorrect.**

* 1. **Table Crime\_Incident**

Number of tuples: **29281**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Screenshot of Crime\_Incident**

**A screenshot of a computer

AI-generated content may be incorrect.**