INST327 Database Design and Modeling
May 2024
Douglas Gomez, Vishal Dalavayi
Patricia Alday, Salih Awel, Adam Zaidi

Project Diary

Introduction to the Problem Domain and Motivation

Throughout human history, managing crime has been of significant societal importance (as it is something that directly hinders human welfare. Crime management requires constant monitoring as well as swift and efficient decision making in order to be effective. Like virtually any location, Montgomery County, Maryland, faces varied forms of crimes daily, ranging from minor misdemeanors to violent high level crimes. As urbanization, socioeconomic trends, and population changes continue to affect Montgomery county as time passes, the need for new, accurate, and extensive data on these crimes rises astronomically. As a result, having access to a reliable, consistent, robust data structure becomes all the more paramount.

A comprehensive cognizance of the scope, nature, details, and tendencies of crime related incidents can lead to optimized resource allocation and improved preventative measures. A structural approach to documenting and organizing crime related data becomes cardinal. This project focuses on creating a MySQL database to record crime related activity in Montgomery county, Maryland, capturing their temporal and spatial attributes, along with demographic and categorization data.

Database Description
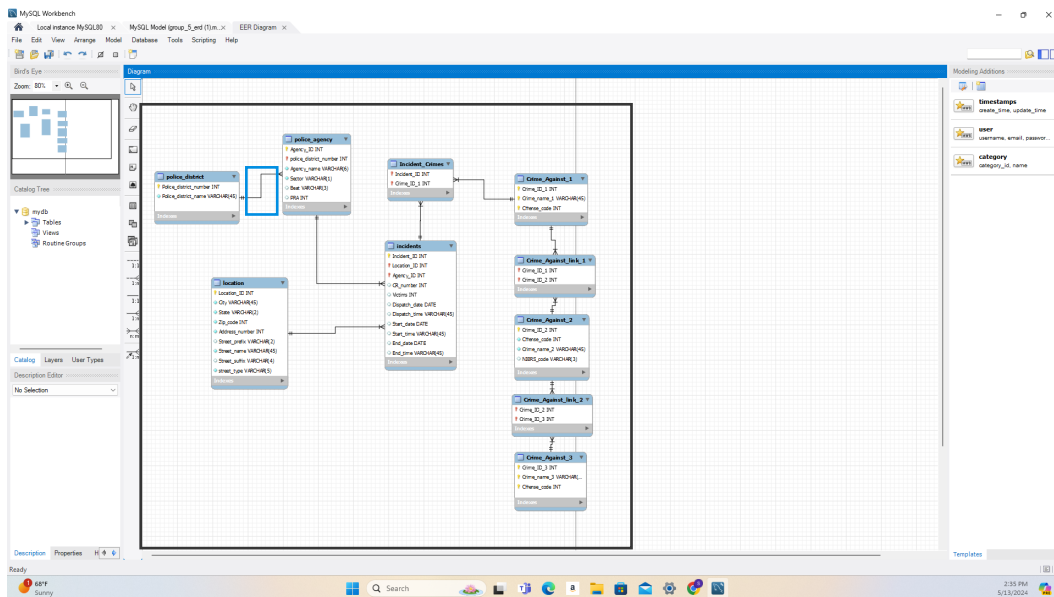
The database is designed to categorize each crime incident based on factors such as location, type of offense, and the responsible law enforcement agency. Key features include:

- Incident Data: Records include Incident IDs (primary keys) linked to specific offense codes, dispatch dates, and location information.
- Crime Classification: The crimes are categorized under various offense codes, distinguishing different types and severities.
- Police Districts and Agencies: Data from police districts and law enforcement agencies provide insights into how each organization handles reported crime incidents.
- Location Information: Captures geographic data, including city, street, and zip code, helping analyze spatial patterns.

With a well-structured schema, the database supports comprehensive queries that can identify crime hotspots, reveal time-based patterns, and assess agency responses. This

functionality empowers policymakers and law enforcement with data-driven insights; and our database views are a reflection of said functionality:

- Locationfilter:
  - Filters information based on specific city = Silver Spring
- IncidentCountByAgency:
  - Creates a view to count the number of incidents by agency
  - Displays the agency names along with their respective incident count
- Crimetrendsovertime:
  - Analyze crime trends over time by aggregating incidents by year and month. a subquery to extract the year and month from the dispatch date.
- CrimeHierarchy:
  - Shows an overview of all the crimes and crime names
- IncidentsbyCity:
  - Incidents in a specific city, filtered using WHERE and the city of Rockville

Database Ethics Considerations

Ethics is always an important consideration to make in any situation. However, it is of even more prevalence in one that deals with crimes of various natures and backgrounds. What can be considered as sensitive information, such as victim details, perpetrator details, and demographic details among other things, should be handled with extreme caution to prevent misuse or discrimination (as well as other outcomes) that are very possible in sensitive situations such as this. Here are some important ethical considerations:

1. Diversity, Equity, and Inclusion: Crime data should not reinforce stereotypes or biases, and queries should avoid targeting specific communities unfairly.

2. Data Privacy: Personal data linked to incidents, such as victim information, should remain secure and accessible only for legitimate purposes.
3. Fair Use and Transparency: Ensure that the data is used responsibly, and provide transparency in how it informs policy decisions.

These considerations guide the database design, ensuring compliance with ethical standards while providing valuable information. By maintaining balance between utility and ethical usage, the database will serve as an important tool for improving public safety in Montgomery County.

Changes from Original Design

After submitting our initial proposal for the database design project, we received valuable feedback from our TA, which prompted us to make significant changes to our design. We revisited our initial list of entities and attributes to realign them with the available data of the Montgomery County crime rates. We focused solely on the entities and attributes that were present in the dataset, rather than making our own– which is what we initially started off with. The team also renamed tables to better describe the data it contained; for example, we renamed the tables that were initially called 'crime_name_1.' Focusing on the ERD, we also reevaluated whether the data types and lengths (e.g., VARCHAR(45)) were appropriate for each attribute based on the expected values. Overall, the most significant change we've made since our project proposal revolved around restructuring our entities and attributes based on the given database to create our ERD model.

We also implemented changes towards our normalization process based on feedback from our progress report. Initially, we had faced issues identified in the 2NF step. We addressed these problems and this involved ensuring that each attribute in our tables was fully functionally dependent on the primary key and that there were no partial dependencies. Our team was able to label and tackle this issue by asking for help from our TA and from there, we worked as a team and were able to fix our 2NF. For our final database, we filtered the data to be within the past 50 crime logs. Filtering the database before importing allowed less rows per table.

Lastly, we were able to improve our communication and collaboration skills. By recognizing the importance of regular communication and collaboration, we were able to establish a schedule for Zoom meetings prior to any major assignments such as the progress reports. These meetings allowed us to discuss and implement effective feedback incorporation and decision-making.

Lessons Learned

Throughout the database design project, our team encountered various challenges and obstacles that provided valuable learning experiences. We learned the importance of remaining flexible and adaptable when working on a complex project like database design. As we encountered issues or discrepancies in the dataset, we had to adjust our approach and make necessary changes to our design– this included reaching out to our TA to help us understand why we got the grade we received or why points were taken off in our assignments. For example, we ran into a misunderstanding when we first created our entities and tables for the database. We learned how to create efficient tables from the entities in the Montgomery County Crime database by the progress report. We learned the value of utilizing available resources, including TAs, peer coaches, module tutorials, and classmates, to overcome challenges and enhance our understanding of database concepts and tools. Not only were we able to learn many technical skills using MySQL, we were also able to acknowledge the importance of collaboration and communication.

Potential Future Work

Reflecting on our database design project, there are several areas where we could implement future improvements and to enhance the functionality, usability, and effectiveness of the database. One potential improvement could be incorporating machine learning models that could predict crime hotspots and pick up on other trends in crime based on current and existing trends. This would be a logical improvement as we already have a huge database of crime that

captured historical data for the machine algorithm to learn off of. Another innovation could be to augment the existing database with a wider range of data points. These data points could include surveillance, community feedback, or socioeconomic variables amongst other things. This would overall provide for a more robust data structure for law agencies to work with. Additionally, we could also improve data security measures in order to protect sensitive information, and keep certain parties anonymous. This would also ensure that the database adheres to ethical standards and security standards simultaneously.

CRUD Queries

For one of our queries, we used the wildcard (*) in order to count all the rows for each distinct value. This was necessary for the query because we wanted to count the number of incidents for each type of crime– regardless of any specific column values.

View Name	Requirement A	Requirement B	Requirement C	Requirement D	Requirement E
locationfilter	X	X			
incidentcountbyagency	X		X	X	
crimetrendsovertime	X	X			X
crimehierarchy	X		X		
incidentsbycity	X	X			

INST327 Database Design and Modeling

May 2024

Douglas Gomez, Vishal Dalavayi

Patricia Alday, Salih Awel, Adam Zaidi

Citations

Montgomery County, MD. "Crime: Open Data Portal." Crime | Open Data Portal, 25 Feb. 2024,

data.montgomerycountymd.gov/Public-Safety/Crime/icn6-v9z3/about_data.