Team 7: Alexia Early, Bryce Daneker, Lily Dinh, Kosy Egbe, Michael Horace

INST327-0102

Final Project Report

20 April 2023

Team 7 Final Project Report

Introduction

The topic our group will focus on is the crime in Montgomery county. We found this topic to be the most interesting option since not only is it close to the University of Maryland, but many students are from Montgomery county or Prince Georges's county, where the university resides. Crime rate is a serious topic as it directly impacts public safety, which is a crucial concern for every community. High crime rates can also lead to increased fear and anxiety among residents, negatively impacting individuals' quality of life. With Montgomery neighboring Prince George's county (which was the previous richest black county in the nation, and now resides in charles county), it's important to research and address the issues within the area that also affect the county's thriving economy. We want the community to be aware of criminal trends and to be able to make informed decisions on safety. Thus, we hope to bring valuable insight to the public about Montgomery County.

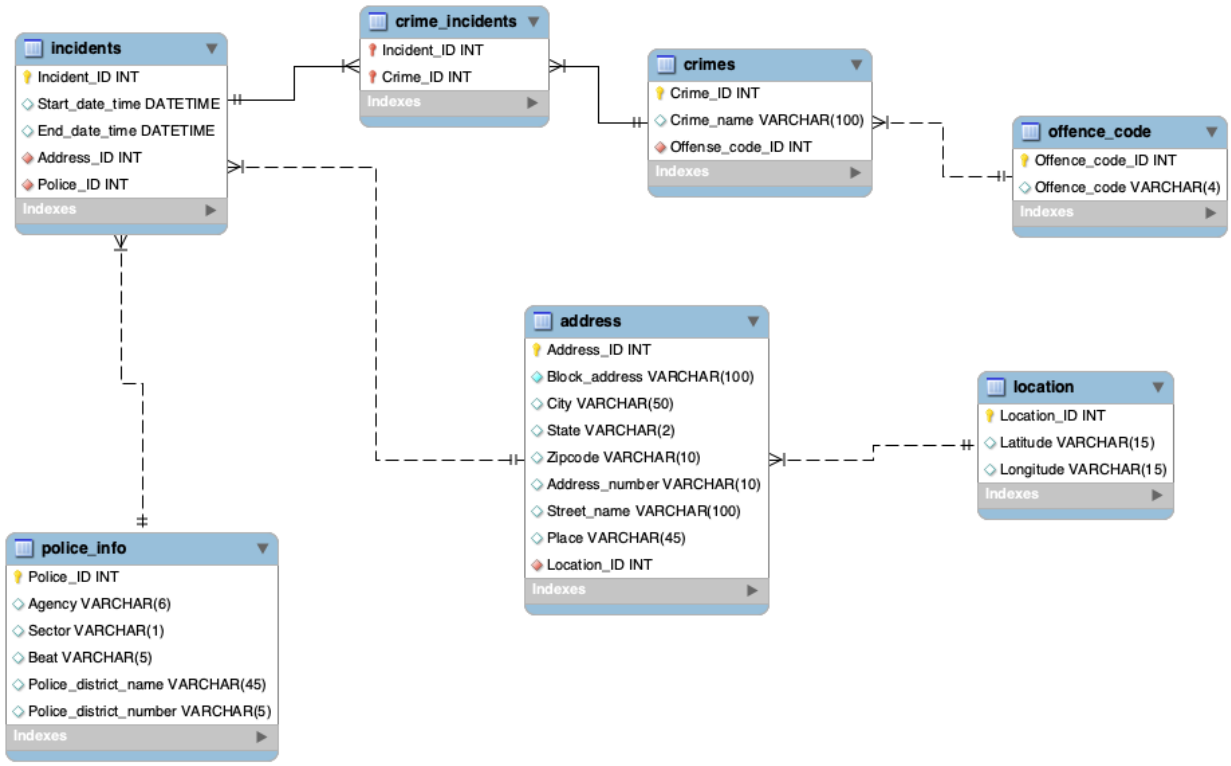
Database Description

The database we designed provides insight to crime data in this area in the Montgomery county region. Our data cover the types of crime, where the crime happened, when the crime happened, and information on the police who responded and reported the crime. We have over 800 entries for the timespan we selected, which was one week in August of 2022. This is to stay within the bounds of the project but to also have an ample amount of data to work with. The

attributes residing will have varying categories of data with characters, integers, and datetime data types.

Logical Design

The figure below shows our ERD for our database. We have seven tables: incidents, police_info, crime_incidents, crimes, offence_code, address, and location. All of our tables rely on one-to-many relationships with the exception of the crime_incidents table, which is a linking table between incidents and crimes. This is because the same incident can have multiple crimes committed. The tables all funnel into the main table, incidents, in order to provide context to the crimes we are discussing. Some locations may have the same coordinates, but similar street addresses which was the purpose of the location table. The same can apply to crimes; some similar crimes such as selling or using marijuana has the same offence code, which is why we have the one-to-many relationship to the offence_code table.



Views / Queries

View Name	Req. A	Req. B	Req. C	Req. D	Req. E
district_crime_date	X	X		X	
no_of_crimes	X		X		X
police_crime_rate	X		X		
silver_spring_crimes	X	X			
specific_crime	X	X			X

Changes From Original Design

The initial focus for our project was on crimes that occurred during the summer and we were going to filter our dataset by the months of July, August, and September, starting from July 16, 2016, when the dataset was first compiled. But as we progressed we refined the scope of our data for the Montgomery County Crime dataset and decided to focus on crimes that occurred only in one week of August. We filtered our dataset by the months oAugus starting from August 15, 2022. This selection was still a reasonable time frame to analyze crime rates during summer and to also ensure we had less than 1000 rows of data for our project. Our focus and potential questions are still within the scope of our data and how we pivot the data to fulfill our needs.

We also made changes to our ERD based on the feedback we received from our AMP and spent time incorporating the suggested changes into the new ERD. The incident ID was removed as the primary key (PK) in the Police Info Table, and it was replaced with the police ID as the sole PK. The Police Incidents Table was deemed unnecessary and was removed from the ERD. Instead, the Police ID was included as a foreign key (FK) in the Incidents Table. The team decided to eliminate the Coordinates and crime code Tables since the necessary information already exists in the Location and Address Tables. This avoids redundant data and simplifies the

database structure. To address the potential table shortage, the team considered creating an offense code Table that can be linked as a primary key (PK) in the crime Table to represent the offense code associated with each crime. The team also acknowledged that there were too many primary keys (PKs) in the Address Table. As a result, we chose to have only one PK, which is the Address ID. Additionally, the primary key location_id from the Location Table was added as a foreign key (FK) in the Address Table to establish the relationship between the two tables.

The primary keys in the Crime Table were modified to include "crime_id" as a primary key which serves as a linking table with the crime_incidents table. Redundant columns in the Crime Table, which were already mentioned, were removed to avoid duplication and maintain data integrity.

Database Ethics Considerations

Based on the limitation of having 1000 rows or less. We had to change our initial focus from summer to one week of August starting from the 1st to 15th. By narrowing the crimes to one specific week in August there is a possibility of excluding certain types of crimes being committed and demographic groups which could lead to potential bias. Along with potentially excluding important trends such as identifying what specific crimes are most committed in Montgomery County. Therefore, the timeframe may not accurately represent the overall most committed crime. Another ethical concern would be inclusion and equity regarding unequal representation of demographic groups by solely focusing on crimes enacted the first week of August.

There potentially can be legal concerns such as liability for the accuracy of the data being used. Ethical concerns can put into question the stigmatization of certain areas or neighborhoods. Unfortunately, while trying to reduce the bias, due to outside factors it may still show bias

against certain racial groups and areas. Outside factors such as over-policing, potential police biases, and population density can play a big role in how our data is created and how it could potentially be digested. Our goal however was to display the data in a way that is factual, least biased, and easily digestible to anyone accessing the dataset.

A big concern we took while working with this dataset was privacy. We felt it is very important to ensure that personal information such as names, address and other identifying characteristics were kept private. It is important to keep personal information private when working on datasets like crime because this information is very sensitive and if it were public it could lead to numerous problems. Some of these problems could include: legal action, loss of public trust, and damage to reputation. Also the individuals would not be able to properly consent to revealing their information for dataset analytic purposes.

Lessons Learned

Throughout the process of finishing this project, the group definitely encountered various challenges and obstacles that required problem-solving and critical thinking skills. One of the biggest challenges was figuring out the normalization of our tables and structure of SQL queries. Normalization was a topic that many of us struggled learning, so we spent ample time with the TAs figuring out the best means of action in regards to getting our tables through the different stages (1NF, 2NF, 3NF). In regards to queries, there were instances where I would spend hours troubleshooting a query that was not returning the expected results, only to realize that there was a minor syntax error in the code and wasn't utilizing the right table to link the different tables.

In some cases, we were able to resolve the issue on our own through research and trial and error, while in other cases, we had to seek help from the instructional team or fellow

members. One instance where some individuals had to seek help was when someone was trying to join multiple tables and was unsure of the correct order of the join clauses.

Another challenge we faced was determining the appropriate data types and constraints to use when creating tables. We learned that selecting the correct data type is critical for ensuring data accuracy and consistency. At times, we had to weigh the pros and cons of different data types and constraints to determine the best option for my specific use case. Also when revamping our ERD model, the incident ID was removed as the primary key (PK) in the Police Info Table, and it was replaced with the police ID as the sole PK. Being able to actively review the work we've and improve it as we go along is an essential skill set that translates into the professional world where even specialists still make mistakes and constantly have to revise and change certain aspects of a working project.

We have learned the importance of attention to detail and careful planning when creating databases and writing queries. Additionally, we have gained a better understanding of the importance of collaboration and seeking help when needed. By utilizing resources such as peers, readings, and the instructional team, we were able to overcome obstacles and create exceptional databases and queries while gaining beneficial skills that can carry into the professional world.

Potential Future Work

Given the opportunity to further our understanding of the dataset and corresponding database, we would expand our scope to include a wider range of data. One of our current constraints was the date range, so in future endeavors, we would widen the range to an entire year to get a better representation of the crimes reported in Montgomery county. We would also like to consider victims of the reported crimes as well as any other information such as number of suspects, or number of felons legally convicted of said crime. This expansion of the current

dataset would give us insight to the crimes being committed in Montgomery county. We could explore questions discussing how many people are on average involved in a crime or if it is likely that there are victims of crimes committed.

Another possible exploration to this dataset would be to add a court_verdict column. This column would detail jail sentencing, if any. This would be beneficial to add because it would allow you to see patterns of sentencing in Montgomery County to certain types of crimes. It could also be used to see how many reported crimes get resolved or not. There are many different ways our team can expand on our current research.