INST 327 - Database Design & Modeling Section 103
Progress Report - 05/03/2024
Team 3
Israel Balbalec II, Ebulo Mejia Jr, Alex Kasaskeris, Hadiya Mohammed, AKM Alam

Final Report

Introduction

Our database will focus on crime in Montgomery County. We plan to create a multi-table database that includes information regarding crimes committed in Montgomery County. The dataset we're sampling includes crimes from July 1, 2016, to January 10, 2023. Our reasoning for choosing this topic is that we believe having a database regarding Montgomery County crime would benefit various groups of people. Many of us live in this area or regularly visit Montgomery County. By creating this database we provide citizens and visitors of Montgomery County peace of mind. They know what areas in the county are safer than others and what areas they should be cautious traversing. We can provide local civil rights organizations and official lawmakers with insight into the issues certain cities or communities are struggling with. This can lead to more discussions involving introducing new policies, laws, or regulations that would make Montgomery County safer. Moreover, we wanted law enforcement to have a database, they can use to determine which areas need more patrolling and how they should prepare certain police units due to some areas having more violent crimes.

Database Description

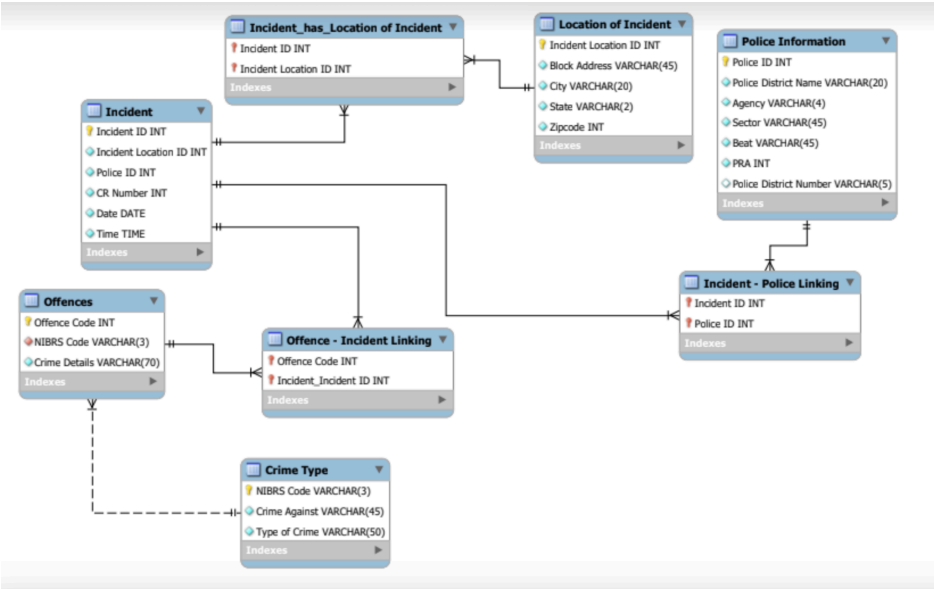
Our database includes 87 rows of the most recent crimes in our provided dataset. Our database contains 7 tables. In our incidents table, we have the incident location ID, police ID, CR number, the date of the crime, and the time of the crime. We determined our incidents had a one-to-many relationship with the location of the crime and what law enforcement group handled the crime. We used auto-incremented foreign keys to show this relationship. In our location of crime table, we used incident location ID as our primary key. The table includes important location data such as the address of the crime, the city, the state, and the zip code. We wanted people who would use our database to know where crimes occurred. In our police information table, we used the police ID as the primary key. For law enforcement to properly use this database they must know who handled the crime in case they need to ask further questions regarding the committed crimes. Our Police information table includes key details about the law enforcement agency that handled each incident. We have the police district's name, the agency,

the sector, the beat, the PRA, and the police district number as attributes in our Police information table.

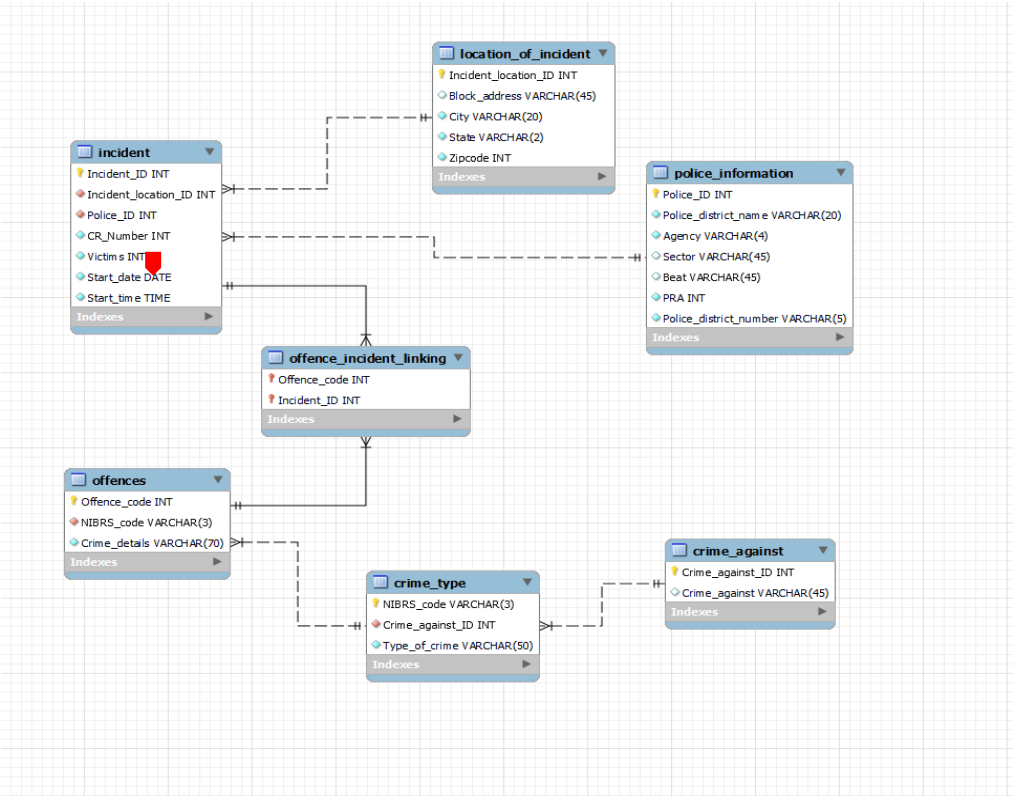
Our next set of tables focuses on detailing what crimes were committed. We noticed that incidents had a many-to-many relationship with offense code, which meant we had to use a linking table to connect our incidents table and offenses table. In the offenses table, we used the offence code as the primary key and had the NIBRS code as the foreign key to our crime type table because we noticed there was a one-to-many relationship between offence code and the NIBRS code. Based on Montgomery County's open data website, crime details are connected to the offence code, which is why made it an attribute in our offences table. For our final table, crime against, we connected that table to our crime type table using an auto-incremented foreign key called Crime against ID to show the one-to-many relationship between the crime_type table to the crime-against table. Similar to how crime details are linked to the offence code, our NIBRS code in the crime type table is linked to the type of crime that was commited.

Logical Design/Physical Database

Initial ERD from Progress Report



Final ERD



Changes from the Original Design

After receiving feedback regarding our progress report's ERD we removed the incident-linking location of the incident table and the incident-linking police information table. We removed these tables because linking tables indicate a many-to-many relationship between two tables. We realized our mistake because our incident table does not have a many-to-many relationship with either the location of the incident table or the police information table. Our incident table has a one-to-many relationship to both the location of the incident and the police agency responsible for handling the incident. We followed Derek's recommendation of using the incident_location_ID and police_ID as foreign keys in our incident table. We added victims to our incident table because we realized it is important to keep note of how many lives were impacted in each incident. After removing our incorrect linking-tables we only had 6 tables, which did not meet the 7-table minimum of the project. Based on our previous normalization meeting with Derek we could create another table based on our crime_type table. We created an auto-incrementing foreign key to indicate a one-to-many relationship between our crime type table and our crime against table.

Sample Data:

The data in our database was retrieved from a CSV file containing all Montgomery County crimes since 2016 and is constantly updated. This file was provided by the professors. As a team, we decided that we would use the last 52 rows for all selected columns/tables, which would be the most recent 52 incidents. This information includes Incident ID, Number of victims, location of the incident, police information, etc. In the beginning of our project we included many columns but as we were working on the project and its normalization, we decided it would be best to exclude certain columns like (latitude, longitude, police sector, beat, PRA). We came to the conclusion that these columns were unnecessary because they didn't provide any relevant information to the database users and would just make the database unnecessarily complex. Most of these changes happened when working on our logical design. Below is a sample of data from our incidents table.

Incident_ID	Incident_loca...	Police_ID	CR_Number	Victims	Start_date	Start_time
201408215	1	1	230001427	1	2023-01-10	00:08:00
201408218	2	3	230001430	1	2023-01-10	00:24:00
201408219	3	2	230001437	1	2023-01-09	18:00:00
201408220	4	4	230001438	1	2023-01-10	06:28:00
201408221	5	5	230001442	1	2022-12-20	00:00:00
201408222	6	6	230001439	1	2023-01-09	19:30:00
201408223	7	7	230001440	1	2023-01-09	19:00:00
201408225	8	5	230001288	1	2023-01-08	18:00:00
201408226	9	8	230001441	3	2023-01-10	06:41:00
201408228	10	9	230001444	1	2023-01-10	08:20:00
201408229	11	10	220057835	1	2022-12-19	09:14:00
201408230	12	11	230001445	1	2023-01-09	19:30:00
201408231	13	12	230001348	1	2023-01-09	14:40:00
201408232	14	13	230001443	1	2022-12-27	00:00:00
201408234	15	14	230001450	1	2023-01-09	22:30:00
201408236	16	7	230001454	1	2023-01-09	12:00:00
201408239	17	15	230001332	1	2023-01-09	12:12:00
201408240	18	12	230001459	1	2023-01-10	09:07:00
201408241	19	8	220054890	1	2022-12-08	10:00:00
201408242	20	12	230001305	1	2023-01-09	10:08:00

This small sample of the incidents table displays incident details.

Views

View Name	Req. A	Req. B	Req. C	Req. D	Req E.
incident_with_details	x	x			
crime_count_by_district	x		x	x	
top_3_crimes	x		x		
crime_aganst_property_by_zipcode	x	x	x	x	x

all_larency_crimes		x			
highest_victim_count	x	x			

Database Ethics Considerations:

Our database consists of opened, closed, reported, and attempted cases that happened in Montgomery County. This means we do not filter or sort any data, but rather take any and all cases that are documented into our database. This eliminates any sort of bias, at least, for incoming data. Our database also does not use any racial profiling or even names so there are no issues with disproportionate targeting. We also do not sort our data to certain areas, we take in all documented crimes in Montgomery County. In our dataset, there are no names of either perpetrator, victim, or the cop attached to the case. We kept all this in mind when creating our database to ensure that, at the very least, it is difficult to know the people involved. We do however have the street police department, city, etc. However, going back to it, it would be very difficult to identify anyone through these statistics alone. Lastly, we are using data that is displayed and shared publicly for anyone anywhere in the U.S. to access and we are in no way modifying or adding on to the data, to ensure fair use. We have carefully conducted our research and created a database, simply for information on reported crimes, without infringing upon anyone's privacy, including the criminals.

Lessons Learned

The project was a great opportunity to work on group dynamics. Our only major roadblock was normalization. Our project's logical design grade was lower than what we expected and that was all due to our normalization. Without proper normalization, we did not feel comfortable moving on to creating our ERDs. We decided to ensure our normalization was close to perfect so our draft ERD in our progress report is was close enough to our final ERD because we didn't want to spend so much of our time to fixing our ERD. Having to adjust and rework our ERD everytime we made a mistake on normalizaing our columns was not optimal. To resolve our lack of understanding regarding normalization, we attended office hours with Derek. We had a decent grasp on normalization, but applying it to our dataset was a struggle for us. Based on Derrick's help, we fixed our normalization. When creating doing the normalization for our project logical design we mistakenly addressed transitive dependencies in 1st NF and created tables intended for

1NF in 3NF. After fully understanding how to properly normalize our dataset we felt more than confident in creating our ERD. We had a small hiccup with our initial ERD in our progress report. We incorrectly labeled the relationship between our incident's table and location of incident's table and our incident's table and police information table. We quickly resolved this problem by creating auto-incrementing FK's for both tables to signify a one-to-many relationship. Our third problem was due to our incorrect ERD in our progress report. We began importing data into our database based on our initial ERD. The importing of our sample data was simple because Trinity recommended to make separate Google sheets for each table to make importing easy. However, we did have to redo a couple of sheets because of our 2 incorrect many-to-many relationships. Fortunately, that didn't take much time and we had our final database after fixing a few tables. Other than those issues the process of creating our database was a positive learning experience.

Potential Future Work

We understand that certain significant factors must be taken into account in order to further enhance our database. To keep the database accurate and functional over time, it is imperative to first ensure data integrity, normalization, and optimization. Regular audits and updates to validate the data and refine the schema to minimize redundancy and data anomalies will contribute to its reliability. Additionally, enhancing the user experience through the development of a user-friendly interface can make accessing and querying the database more intuitive for all stakeholders. Incorporating data visualization tools will facilitate the interpretation of complex crime data, allowing users to identify trends and patterns more easily. Real-time updates would provide users with the most current information available, ensuring that decisions are based on the latest data.

Integrating external data sources, such as demographic or socioeconomic data, would enrich the database, providing additional context for understanding crime patterns. Furthermore, leveraging predictive analytics techniques could enable us to forecast future crime trends based on historical data, empowering law enforcement agencies and policymakers to proactively allocate resources and implement preventive measures.

Continued collaboration with law enforcement agencies, community organizations, and policymakers, is essential to ensure that the database remains relevant and meets the evolving needs of Montgomery County's residents, visitors, and law enforcement agencies. Moreover,

maintaining a focus on privacy and security measures is paramount to protect sensitive information and comply with data protection regulations, thereby fostering trust and confidence in the database among its users. By addressing these considerations and continuously iterating on the database, we aim to enhance its value and utility in addressing the complex challenges of crime in Montgomery County.

Citation for dataset

Montgomery County, MD. (2015a, March 24). Crime: Open data portal. Crime | Open Data Portal. https://data.montgomerycountymd.gov/Public-Safety/Crime/icn6-v9z3/about_data