INST 327, Section 0103

Project Report

05/12/24

Team 5: Jordan Ades, Nicola Munuo, Lynell Nchako, Joan Ojukwu, James Swift

Introduction:

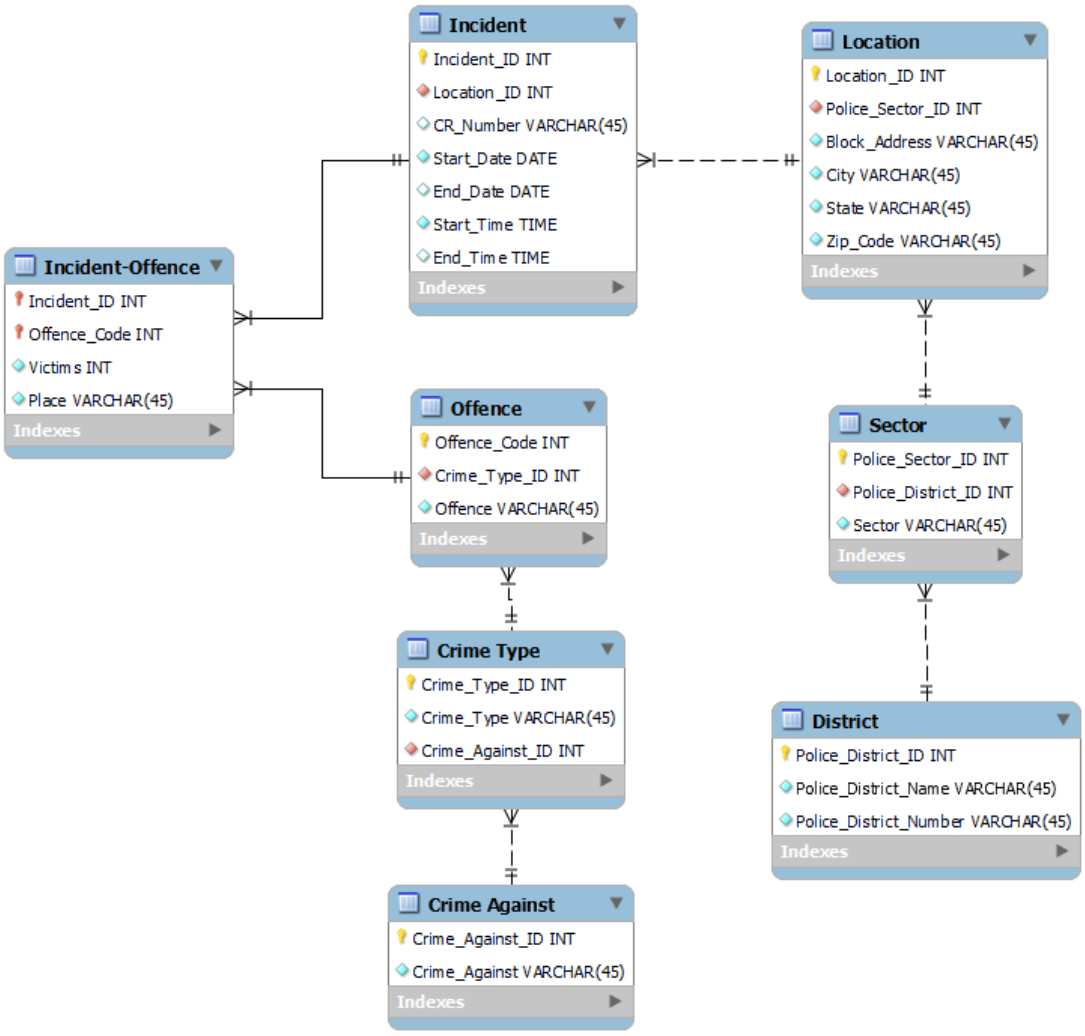
The database our team is using contains data on the crime statistics in Montgomery County Maryland. It has a wide array of crimes, even including open cases and reported crimes. The original dataset we were given contained an excessive amount of data that we did not fully need for the basis of this project, so we trimmed it down to only include some of the more general and important pieces of information, such as the location and time of the incident, and more. We decided to do this simplification to help make the data easier for users to navigate, as our plan is for it to be used to help people who live in or are planning to live in Montgomery County learn how safe/dangerous certain areas are. While the data might not be the most clear reflection of how safe a certain place actually is, knowing how many crimes happen there is helpful for getting a general idea.

We believe this could be very helpful to those previously mentioned, and we also have high faith in the data as it was obtained from the Montgomery County government. Knowing specific crimes in certain areas would also not just be helpful for residents, but also for business owners, for example, someone would not want to open a store in a neighborhood with a high burglary rate. We hope that our database can be used to assist the general public, or even just educate people on what goes on around them.

Database Description:

Our database is constructed around various data elements concerning crime in Montgomery County, Maryland. It is meant to aid those looking to learn more about crime statistics and those looking to learn more about trends and patterns for crime in Montgomery County. Our database contains 8 tables and encompasses multiple elements of crime data and information.

Logical Design:



Our main goal when creating a logical design was to create a database that is easy to navigate and allows users to fully take advantage of its functionality. We created our database to have eight tables because that number allowed the database to be easily adapted and manipulated without it becoming too unnecessarily complicated.

Our first main discovery when producing our logical design was how many elements repeated in our dataset. For example, ‘incident_id’ had many repeat values. This allowed us to deduce that for any one incident, multiple crimes could be committed. And, of course, one crime could be committed in various incidents. This allowed us to see that we needed to have an incident table with ‘incident_id’ as a primary key and an offense table with ‘offence_code’ as its primary key. We then created a linking table ‘Incident-Offense’ that contains both ‘incident_id’

and ‘offence_code’ as foreign/composite keys. Once we had these three tables, we were easily able to then further normalize the structure of the database. We broke down other elements into multiple tables to reduce repeating values and increase the functionality of the database as a whole.

When determining the columns that we wanted to accept non-null values, we first analyzed the dataset itself to see what values it always included and which values were optional. For example, the ‘End_Date’ column in the ‘Incident’ table can have null values. This is because in the dataset, there was not always a listed end date/time for an incident. With this in mind, we decided that this value could be null in our database.

Physical Database:

There were not any substantial changes to our design in the process between the logical design and the physical database creation. We took our sample data from the dataset given and organized it into the tables that we designed. During this process, we assigned values to multiple primary keys that did not exist in the original dataset. Columns like ‘location_id’ and ‘crime_type_id’ were created in order to better organize the database so these values had to be assigned for each of their corresponding values. Once all of the tables were filled, they were easily able to be imported into the database.

Sample Data:

We took our sample data from the large dataset that we were given to work with. Our dataset contained various crime data from all around Montgomery County, MD. Given this, we wanted to make sure that the sample data we used contained various crimes, incidents, and locations. This is to make sure that each table’s functionality was easily apparent to any users who weren’t familiar with the dataset. Organizing the dataset by ‘incident_id’ and then selecting a random subset of this data gave us a snapshot that contained various values for crime and location data. Using a random subset of data best displays what the database would look like if

we had used all of the given data. Below is a section of our 'incident' table.

Incident_ID	Location_ID	CR_Number	Start_Date	End_Date	Start_Time	End_Time
201087096	1	16033231	2016-07-01	NULL	00:22:00	NULL
201087097	2	16033232	2016-07-01	NULL	00:04:00	NULL
201087100	3	16033233	2016-07-01	NULL	00:32:00	NULL
201087102	4	16033238	2016-07-01	2016-07-01	00:58:00	00:58:00
201087104	5	16033235	2016-07-01	2016-07-01	00:57:00	01:00:00
201087108	6	16033248	2016-07-01	NULL	04:00:00	NULL
201087119	7	16033252	2016-07-01	2016-07-01	05:00:00	05:30:00
201087133	8	16033268	2016-07-01	NULL	08:52:00	NULL
201087135	9	16033263	2016-07-01	NULL	08:15:00	NULL

There are some tables that were not able to have 30 rows. For tables like 'district', 'crime against', 'crime type', etc., there are not enough types of these to have 30 rows. For these, the number of rows is the number of unique primary key values for that given table.

Views/Queries:

	Query Name	**Req A**	**Req B**	**Req C**	**Req D**	**Req E**
#1	incidents_overview	X				
#2	open_cases	X	X			
#3	multiple_incidents	X	X	X		
#4	209_zipcode	X	X			X
#5	substance_related	X	X		X	
#6	incidents_per_victim_type	X		X	X	
#7	res_incidents_per_district	X	X	X		X

Query Descriptions:

1. Creates a view called incidents_overview and shows a description of each incident, including its place and crime descriptions, ordered by the victim type and then the crime type
2. Creates a view called open_cases that shows all of the currently open police reports that have a start date of 7/1/2016. Its rows are sorted by sector and city, respectively.
3. Creates a view called multiple_incidents that counts number of incidents that occured in each city on 7/1/2016, if the count is above 1.
4. Creates a view called 209_zipcode that shows both open and closed police reports, that have zip codes beginning with '209'.
5. Creates a view called substance_related query that shows the incidents that involved drugs or alcohol in public places.
6. Creates a view called incident_per_victim_type that counts the number of incidents for each victim type (people, property, and society).
7. Creates a view called res_incidents_per_district that counts the number of incidents that took place in residential areas, for each police district.

Changes from the Original Design:

Throughout the semester, we received feedback from the instructional team, which helped us to specify the overall theme of our inquiries with the sample data. We were intentional in our use of primary keys and composite primary keys in our tables, and created appropriate linking tables with many-to-many relationships. For example, for our dataset, we ensured that an incident could not occur in various locations, while a location could have various incidents. We also updated our plan for the sample data, as the original dataset contained repetitive or irrelevant data that we aimed to exclude. Specifically, we chose not to include the NIBRS code, Crime Name2 (which describes the NIBRS_CODE), the geolocation of a crime (latitude and longitude), or any data that is a subset of other data. Our original design aimed to answer questions such as "How many officers are typically needed to deal with different types of crimes?" or "How many officers do different police departments typically dispatch?" However, upon further reflection on the scope of our design, we discovered that the database cannot provide that amount of police officers for any of the data elements. Furthermore, while we still aim to consider diversity, equity, and inclusion in our design, we were unable to summarize the demographic composition of the communities affected by high rates of crime. This would have enabled a nuanced

understanding of how different parts of the population in Montgomery County are impacted by the rates of crime**.**

Database Ethics Considerations:

We made sure to consider ethical concerns when designing our database, particularly principles of data privacy, fair use, diversity, equity, and inclusion. We believe that it is crucial to include demographic diversity in our database, to provide a comprehensive view of crime across Montgomery County. Our database covers a range of crimes and locations within the county, aiming to eliminate biases and promote transparency in law enforcement and policymaking. Although we were unable to collect demographic data, we believe that the locations of crimes committed and their density can still convey important information. To protect data privacy, we avoided using identifiable information for any victims, witnesses, or suspects, instead of using counts to track the number of each group in our database. We recognize the risks associated with specifying the exact location of a crime, such as street addresses or building numbers, and therefore, we have opted to use only cities and zip codes for location-based inquiries.

Rawls' theory of justice, particularly the concept of the "veil of ignorance", is something we also considered in our approach to this database. The veil asks us to imagine ourselves as unaware of our social status, wealth, or other characteristics when designing principles of justice. We aimed to cover the interests and needs of all individuals within and hope to live in Montgomery County, regardless of their demographic characteristics. This approach can help ensure fairness and equity in the representation of crime data. We also incorporated Derrida's ethics of writing and responsibility in our approach to data privacy. In designing our database, prioritizing the protection of individuals' privacy rights aligns with Derrida's emphasis on respect for the Other (represents the difference of existence beyond the boundaries of the self) and the avoidance of harm. Using the Rawlsian perspective for concerns about data privacy, we can also relate to the principle of equal basic liberties. Rawls argues that individuals have a right to privacy as part of their basic liberties, which should be protected in the design of social institutions. Ensuring that the database avoids using identifiable information for victims, witnesses, and suspects aligns with this principle, as it respects individuals' rights to privacy and autonomy.

Lessons Learned:

Throughout the duration of this project, our team acquired invaluable insights into various aspects of database design and development, which will undoubtedly influence our future endeavors. One of the most crucial takeaways was the significance of adhering to strict normalization steps (1NF, 2NF, 3NF) to maintain data integrity, reduce redundancy, and mitigate issues like partial dependencies and repeating data. We realized that sticking to these principles is vital for sustaining a well-organized and efficient database. Additionally, we deepened our understanding of modeling one-to-many and many-to-many relationships between entities using primary keys and foreign keys. Accurately representing these relationships is pivotal for ensuring data consistency and enabling robust querying and reporting capabilities. Another significant lesson was recognizing the importance of establishing a clear scope and sampling plan for the data to be included, rather than incorporating all available data indiscriminately. This approach not only enhances database performance but also ensures that the analyzed data is relevant to the project's objectives. Moreover, we heightened our awareness of considering diversity, equity, inclusion, and ethical perspectives, such as data privacy, when working with public data sources. As database professionals, it is our responsibility to ensure that our designs and implementations do not perpetuate biases or compromise the privacy and security of sensitive information. Finally, we learned the importance of paying attention to details such as proper header/footer formatting, in-text citations, and meaningful file naming conventions for project submissions, as these seemingly minor aspects significantly contribute to the overall professionalism and quality of our work.

Potential Future Work:

As we move ahead, one area we're focusing on is improving our database setup based on the helpful feedback we got about normalization and properly organizing how different things in the database relate to each other. We're planning to do a thorough check to make sure all the tables have clear main keys, and if needed, combined keys, and the right connections to keep everything linked properly across the database. We're also thinking about only including location data that are tied to one police sector to avoid making things too complicated with how locations and sectors relate. Also, we understand how important it is to give a detailed introduction and description of our database, especially explaining the specific details of the crime data. We'll clearly explain why we chose this topic and show why analyzing this data matters. We're also

going to look into ways to make our database more diverse by including information about different demographics and making sure we're not overlooking any biases or having unequal impacts. This work fits with our goal of making sure diversity, fairness, and inclusion are part of every step of creating and using the database.

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.

Freeman, Samuel. "Original Position (Stanford Encyclopedia of Philosophy)." *Stanford Encyclopedia of Philosophy*, 20 December 2008, https://plato.stanford.edu/entries/original-position/. Accessed 11 May 2024.

Anderson, Nicole. "Derrida: Ethics Under Erasure (London, Bloomsbury, 2012)." Edinburgh University Press, Edinburgh University Press, 8 January 2022, https://www.euppublishing.com/doi/full/10.3366/olr.2015.0172. Accessed 11 May 2024.