INST 327, Section: 0102
Project Final Report - 12/14/2023
Team 5: Brandon Kang, Noah Kandel, Aki Parrish, Patricia Louis, Dylan Fisher

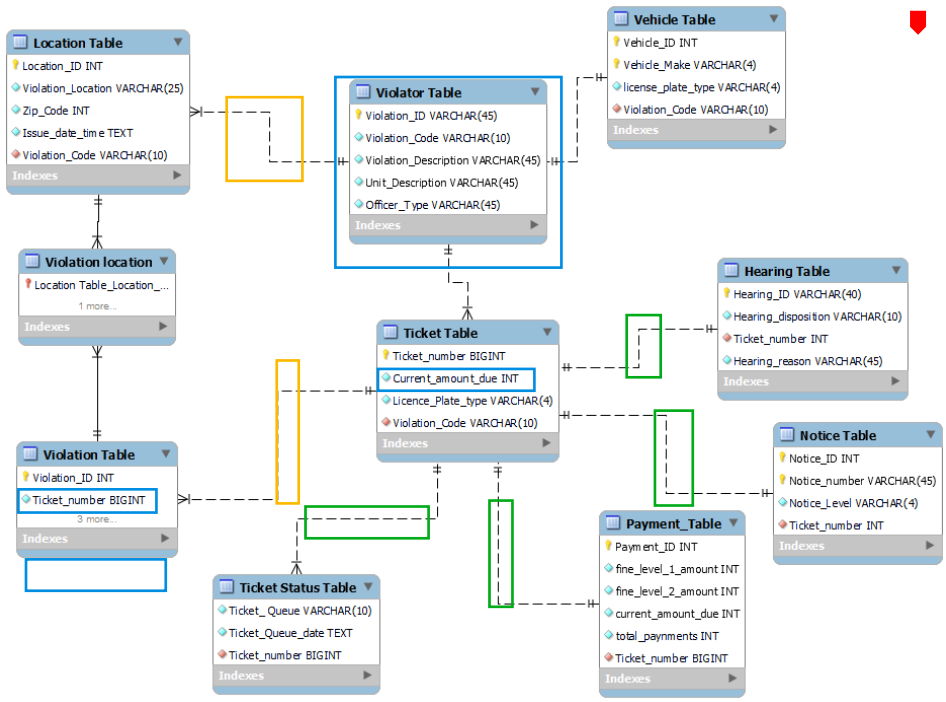
Introduction:

Our collaborative efforts have resulted in the development of a database designed to organize information related to traffic violations in the bustling city of Chicago. This repository serves as a resource for law enforcement, offering valuable insights into the incidents that constitute breaches of motor vehicle operation laws on the city's streets and highways. Our main goal is to empower Chicago law enforcement agencies with the tools necessary to proactively address and mitigate traffic violations, thereby contributing to the broader objective of creating safer thoroughfares for the city's residents. The database we gathered data from is structured to centralize data on a significant scale, encompassing details from 1000 distinct traffic violations. This strategic selection of relevant columns ensures that the database is not only robust but also tailored to the specific needs and priorities of Chicago law enforcement. By focusing on key parameters such as violation type, location, and time, our database provides a comprehensive understanding of the traffic violation landscape in the city. To refine the database and enhance its practicality, we curated a subset of 300 traffic violations from the larger dataset obtained from ProPublica. This subset is derived from "every ticket and warning issued by the City of Chicago through its red-light and speed camera programs since January 1, 2010." This deliberate reduction in dataset volume serves a dual purpose: it facilitates a more manageable and user-friendly database for our team and, more importantly, ensures that the information presented is highly relevant and actionable for law enforcement professionals. In shaping this database, we sought to address specific challenges identified in the original dataset, such as information overload and inconsistent formatting. By carefully selecting and organizing data columns, including violation details, locations, and time stamps, we aimed to create a user-friendly and efficient tool for Chicago law enforcement to navigate and analyze traffic violations effectively.

Database Description:

Our database was constructed to find the connections between different factors on the database of Chicago traffic violations in order to find a pattern in the traffic violations that happened in the past. The objective is to find the pattern in traffic violations in order to prevent violations that could be happening, to help improve safety, and to sort the database into contained tables to access data easier and have less clutter when looking for specific information.

INST 327, Section: 0102
Project Final Report - 12/14/2023
Team 5: Brandon Kang, Noah Kandel, Aki Parrish, Patricia Louis, Dylan Fisher



Logical Design:

When building the logical design, we started by creating many tables that could help find a pattern of traffic violations. As the process continued to brush up the table, we eliminated the table with duplicated data and tidied up the relationship between the tables from having many one-to-one relationships to a few number of one-to-many and many-to-many. Almost everything revolves around the ticket number as the ticket number will always be unique and not have duplicate numbers, which is why many of the tables are connected to the ticket table in the center. We saw that there was related information regarding payments for fines and payments made, so we gathered them into a payment_table, and the same was done for ticket notices and statuses, as we didn't want the main ticket table to be overcrowded with keys. We decided to split violations and locations with a linking table, as the same type of violation can occur in many different locations and one location can have many different violations.

Physical Database:

Our database aims to reduce the amount of unnecessary information when looking at the database for chicago traffic citations by dividing columns into their related tables. Since every

ticket should always have a ticket number, many of the tables are linked through the ticket number as a foreign key. We thought about how violations have a lot of information regarding both the actual details of the violations and the details on the officer's end with officer type and unit descriptions, so we split those into two different tables for violations and violators. We also separated the locations into its own table due to the street names and zip codes being available. Vehicles involved in violations could also be important for identifying trends through the data, such as certain car makes being involved in more violations than others, so we moved the related data (including vehicle make and license plate type) into its own table as well. We originally thought about not including hearing data and notices because of the broad scope of our database, but we included it in the end because there could be possible patterns in whether certain types of tickets are more often taken to court, or paid off based on violation severity.

Sample Data:

Our sample data focuses on traffic violations, centering around the Ticket table. This core table includes ticket numbers, current fines, license plate types, and violation codes linked to related tables. The Financial and Payment tables detail fine amounts and payment patterns, while the Notice and Hearing tables track hearing-related information. The Violation code links to the Violator table, providing violator details. The Vehicle table adds vehicle characteristics, and the Location table lists violation locations. This streamlined structure enables law enforcement to analyze ticket-related trends, allocate resources efficiently, and enhance traffic safety efforts. For example, our ticket table looks like this:

ticket_number	current_amount_due	license_plate_type	violation_code
69006797	0	PAS	0964080B
69006798	0	PAS	0964125B
71374497	200	PAS	0964080B
71374498	200	PAS	0964080B
71374499	0	PAS	0964080B
71374500	200	PAS	0964080B
71374522	0	PAS	0964050J
71379641	0	PAS	0964080B
71379642	200	PAS	0964080B
71395908	300	PAS	0964100A
71405808	0	PTR	0964150B
71405809	150	PAS	0964090E
71405810	150	PAS	0964090E
71405811	150	PAS	0964090E
71425436	0	PAS	0964050J

Views and Queries:

View Name	Joins (A)	Filtering (B)	Aggregation (C)	Linking (D)	Subquery (E)
Ford/Nissan Car Tickets	**X**	**X**			
Full Ticket Info	**X**	**X**		**X**	
Most Common Violations	**X**		**X**		
Problematic Locations	**X**		**X**		**X**
Unpaid Ticket Info	**X**	**X**			

Changes From Original Design:

The development of our database took place over several months and within that time several changes were made. These changes were made so that the database would be properly structured and therefore serve as a good source for our intended audience. Many of the changes included removing and adding data attributes to tables so that the database was more organized in terms of navigation and where to retrieve information. For example, 'ticket_number' was added to multiple as a foreign key in order to connect tables. Another major change that was made was converting the tables so that the necessary relationships would be one-to-many instead of one-to-one. In order to do this, we changed the structure of the ERD so that tables would serve as a linking table when needed. We added the necessary foreign keys to do so as well. We also altered our database to accommodate ethical concerns, which is discussed in detail in the next section.

Database Ethics Considerations:

While designing the database it was important to keep the ethicality of it in mind. The database contains information on Chicago citizens that could be considered sensitive and therefore it is our responsibility to ensure that their data is protected. While the dataset itself does a good job at providing anonymity to all parties involved, it is still possible to trace the data back

to the individual. This is especially true considering that the intended viewer, law enforcement agencies, have access to additional information that would let them do just that. Therefore, our team had to make sure that we minimized the amount of data made available. The purpose of this database is to act as a resource for law enforcement to gain a better sense of the traffic violations in the city in order to reduce them. We determined that not every data attribute originally provided was necessary to achieve this. Therefore, we decided to exclude the "Hearing Reason" attribute from the database because this would not have been useful information to the viewer since they are more so focused on the traffic violation itself rather than the aftermath (unless it is known it can have an impact on future violations). Another ethical concern is the possibility of allowing people to view and modify the data. In order to manage this, we included views which will allow viewers to view the data without giving them the ability to change any data. This gives further protection to user data and minimizes the chances of their information being tampered with, which could have severe consequences.

Another aspect of database ethics is ensuring that the diversity, equity and inclusion of the database is considered. The dataset is unbiased in that it covers the entirety of Chicago and every single citation that was recorded into the system. Due to the sheer size of the dataset, our team had to figure out how we would limit it so that it was more manageable in SQL. Considering that our database is meant to represent all people in Chicago, we had to limit the data in a way that would not unfairly exclude any groups of people. We decided to only include the first 300 rows which limits the dataset by the time the tickets were issued (1/1/21 0:00 to 1/1/21 2:48). Since there is no bias between the groups of people out during that time, the dataset still represents all people. The attributes in the provided dataset seemingly also have no biases when reporting the traffic citation data. All the attributes were factual and straightforward, providing no room for biased information to slip in. Therefore, our team had to make no changes to the data attributes when it came to inclusion concerns. While the attributes provided were not biased, the absence of certain attributes could lead to equity concerns. The dataset lacks information about the socioeconomic, racial and gender demographics involved in each citation. The database is made to help law enforcement agencies or closely related organizations track traffic citations and minimize the amount of violations. It can be assumed that these organizations will change their own procedures as well as take their findings to policymakers

who will then make decisions based on the information. When making decisions that will significantly impact a community, it is important to make sure that you have the full picture of said community. Demographic information should be provided so that the viewer is given more context behind the data presented. Without context it is possible that viewers may make a decision that disproportionately impacts a certain group. For example, law enforcement agencies may see that a certain area has a higher rate of traffic violations than others. They will likely then decide to set harsher penalties to discourage people from committing these violations. However, what they missed was that the location of the violation was in a low income area, which tend to be heavily policed. This means that instead of the people being more prone to commit violations like assumed, the reality could be that people are more likely to get caught by police or cameras in the area. However, since the viewer did not have this information, these people are now being unfairly punished by higher penalties that a good amount of them may have trouble with paying. It is because of scenarios like these that it is crucial to have this information available to viewers of the database.

Lessons Learned:

In our collaborative effort to make and design a comprehensive traffic violation database for Chicago law enforcement, we have learned many valuable lessons, helping shape our understanding of data management, ethical considerations with our database, and effective teamwork. The lessons we learned not only helped us to create a successful development of the database but also provided insights for other projects we may end up doing in our jobs or other classes.

One main fundamental lesson learned throughout our project revolves around the process of data selection and refinement of the data. The initial dataset, sourced from ProPublica, was extensive, which required us to only grab relevant data that we could use. By focusing on columns and reducing the dataset to 1000 distinct traffic violations, we addressed challenges related to information overload. This lesson shows the significance of thoughtful data categorization in creating a manageable and actionable resource. The design and organization of the database structure taught us to balance between comprehensiveness and user-friendliness. The careful structuring of data into their distinct tables; ranging from the core Ticket table to Financial, Payment, Notice, Hearing, Violator, Vehicle, and Location tables, proved to be

important for the analysis and resource allocation by law enforcement. This lesson emphasized the importance of tailoring database structure to your needs and priorities of end-users.

Lastly, collaboration emerged as a key aspect of the project, teaching us the significance of effective communication and shared understanding. Regular updates on our project, feedback sessions in GroupMe, and an open mindset were essential for tackling challenges and ensuring the successful development of the database. This lesson shows the need for open communication channels and collaborative efforts in complex projects.

Potential Future Work:

As we reflect on the development of our traffic violation database for Chicago law enforcement, the project can also serve as a launchpad for potential future work. The lessons we learned from this project have opened the door to innovative advancements in our project that could further enhance the effectiveness of our database.

The integration of real-time data emerges as a massive enhancement for our database. By providing law enforcement with up-to-date information on traffic violations, the system could facilitate quicker response times to emergencies and issues. Real-time updates would empower agencies to deploy resources more effectively, responding proactively to changing dynamics on the streets. Another thing we could work on is regular database maintenance and updates are imperative for the sustained relevance of the project. As traffic violation patterns change over time, continuous adjustments to our database structure and new data ensure its ongoing effectiveness. Implementing an automated system for regular updates and maintenance checks could smooth this process, keeping the database current and reflective of the dynamic nature of traffic behavior.

Lastly, collaboration with real city planners can help us strategize our future work. Connecting the database with city officials could improve infrastructure and policy changes in areas with more traffic violations. By aligning our database with law enforcement our project could contribute to a more integrated system with an approach to traffic management in Chicago.

In conclusion, our potential future work for the Chicago traffic violation database project is extensive. However, each advancement represents an opportunity to enhance the database's impact on traffic safety, contributing to safer and more informed roadways for Chicago.