INST327

Section 0203

05/14/2023

Gabriel Bekele, Vincent Chung, Nabiha Athar

Team 8

Team Project Final Report

Introduction

For our project, our team focused on Chicago traffic citations. Seeing as Chicago is the third-largest city in the United States by population, we figured there would be a large dataset of information for us to work from. It is off of this dataset that we developed our ERD model, and pulled data to serve as sample data for our physical database. Overall, we feel as though we picked a good topic to work on because it was interesting to analyze the data, as it consisted of so many records, spanning over a decade. We also believe that the database that has been produced is extremely helpful because it could potentially be used to help the police gather different kinds of information. For example, they will be able to see if there are specific intersections in Chicago that have been causing the most trouble, and perhaps they could put up a stop sign or a light. Overall, this is a dataset that would benefit the police not only within Chicago, but even nationwide.

Database Description

The information in our database includes the date that the ticket was issued, the location and zip code of the violation, a brief description of said violation, the vehicle make and the license plate type. Tickets that were logged in this database were given for a multitude of reasons such as parking violations, speeding, and running red lights since January 1st, 2010. This was a rather large dataset as hundreds of violations are committed every hour thus putting us in a position to narrow down our scope in order to make it more manageable.

Logical Design

We wanted our ERD model to be intuitive and simplistic so that users could understand the tables and relationships without any confusion. We centered our model around the tickets that were given because the ticket is the clear identifier that any violation had occurred and due to the

nature of our sample data the ticket number serves as a great unique identifier, as the number is different for each ticket. This allowed it to be very useful for linking as a primary key of the tickets table and as a foreign and composite primary key for many others. Other aspects of our dataset have repeating values and because of that we were unable to create unique keys making it much more difficult to center an ERD around.

The logical design for our project, while somewhat challenging, was completed rather quickly. We credit this success to having a solid understanding of the concepts being used in the process such as the ethical considerations that had to be made in terms of the scope of the project as well technical aspects like developing an ERD. Some of the issues that we found when looking over our model was that some of our formatting was incorrect, however that was a rather easy fix. We also noticed that some of the variables were missing data types, so we added those in. In addition to having variables that were missing data types we also have improperly assigned ones. One that particularly had us stuck for a while was the ticket number and notice number which we had incorrectly as a VARCHAR at first but then changed it to an INT, which still had problems due to most of the data points exceeding the max value for a variable of the INT data type. We managed to log a lot of information in our ERD as well, including the notices, fine amount, violations, and hearings among many other things. We chose to focus on tickets because it is the feature that is unique to every traffic citation and allows the navigation of our model and database to be easily comprehended. Below is a picture of our ERD with all the updates that were made.
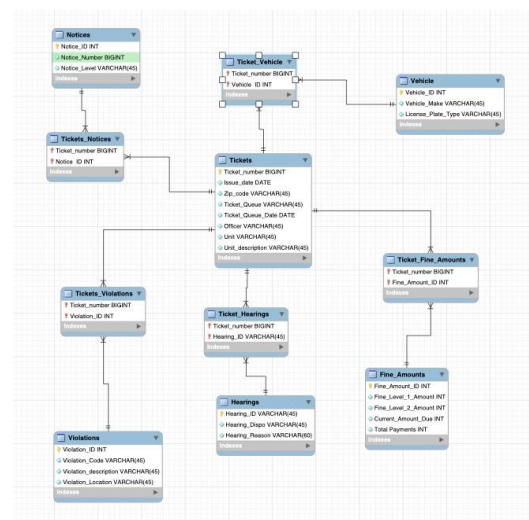
Physical Database

      Our database has a total of eleven tables: notices, tickets_notices, tickets, ticket_vehicle, vehicle, tickets_violations, violations, ticket_hearings, hearings, ticket_fine_amounts, and fine_amounts. In order to link the tables to each other we used a many to many relationship between ticket_number (primary key) and all other tables. We matched ticket_number to all other respective IDs such as Hearing_ID, Vehicle_ID, etc. The linking tables are: tickets_violations, ticket_hearings, ticket_fine_amounts, and ticket_notices, and ticket_vehicle. Below is a screenshot of the tables:
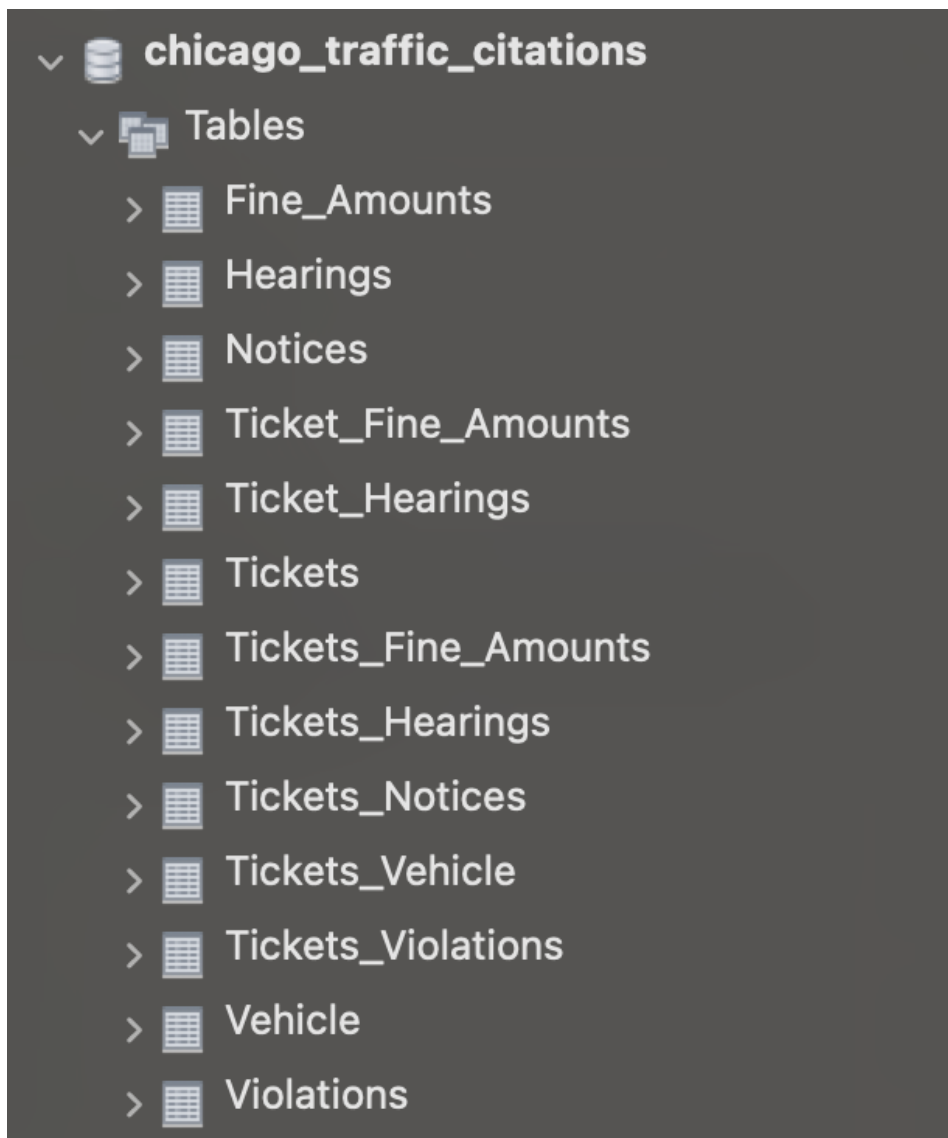
Sample Data

We found a dataset that spanned over thirteen years and had a plethora of data that we could utilize. Due to the dataset including over ten years worth of traffic citation records and there being thousands of citations recorded in a day, the scope of the database was condensed so that it only includes traffic citations recorded on January 1, 2021. Many people travel and have events to attend on New Year's Day so the number of citations is relatively higher compared to an average day, with hundreds of citations being handed out in the first hour of the year 2021. This led us to decide to condense this dataset to the first hour of the New Year and get some perspective on the hectic night that it is. This still gives us a large dataset to work with without it being overbearing or confusing. We used different parts of the dataset such as tickets, violations, and hearings.

Views/Queries

The CRUD operations posed the biggest challenge to us. As a group of three, we all had to work together to complete this accurately. We made five queries that were saved as views about our database. Our queries returned the number of tickets using different filters such as vehicle make and violation codes split by plate types. We saved all our queries as views to make them more accessible since the SQL code is not necessary to access the information.

View Name	Req. A	Req. B	Req. C	Req. D	Req. E
queryone	X		X	X	
querytwo	X	X		X	
querythree	X			X	
queryfour	X	X		X	X
queryfive	X	X		X	

Query One: This query, saved as a view, returns the number of tickets per vehicle make

Query Two: This query, saved as a view, returns the number of tickets per violation code split by plate types

Query Three: This query, saved as a view, returns the average amount paid based on police unit

Query Four: This query, saved as a view, returns the tickets have money owed

Query Five: This query, saved as a view, returns the tickets where the notice level is 'VIOL'

Changes From Original Design

The most significant change we made while working on this project was picking a topic and sticking to it. We had initially chosen to work on Chicago traffic citations as well as the electric vehicle population in the United States. After receiving feedback from the TA's we chose to work on the Chicago traffic citations instead of the EV population because ultimately, we felt that this would be more interesting and have more data for us to work with. We then switched from Chicago traffic citations to Montgomery County traffic citations. We originally wanted to do this switch because we felt as though it was more relatable to us as Maryland residents. We changed our entire proposal to focus on Montgomery County traffic citations before meeting with a TA and realizing that we were unable to do this. We then switched back to Chicago traffic citations and revised our original project proposal. Other than the switching of topics we did not change much from our original design. We edited our ERD and took out a table or two to make it neater and more precise. Ultimately, we feel as though we took our time with this project and didn't need to change much because of the care and accuracy we initially put in.

Database Ethics Considerations

When considering the database ethics, it's hard to apply some of the readings and their implications onto our dataset. The information that was given to us did not provide us with details such as race or gender of the people that were stopped, so there is no way to see if there were any stereotypes or biases at play. There is also no AI that is being utilized so we are unable to apply the ethics of that onto our database either.

Lessons Learned

Some of the biggest lessons learned from working on this project came when we were trying to import our data into our database. We realized that our database name could not have

any capital letters, and as a result, we had to edit our database creation script and re-import the entirety of our data. We also had to update our ERD every time we made a change to the database script and vice versa. For example, if we changed a datatype in our script, we would have to go back and update the ERD to reflect this change. This ended up being a very time-consuming process, but in the end we managed to make it work.

Potential Future Work

If we had more time to work on this project, we would have tried to manage a bigger scope. By having a bigger scope we would have more data to analyze and compare. This would give us a better understanding of the traffic citations and patterns in Chicago. We also could have found a more comprehensive data set with more information. For example, if we had found a dataset that also included the race or gender of the driver the ticket was issued to, we could analyze whether there were any ethical considerations. We also could have added more queries which would have further narrowed down our database according to the results we wanted.