INST 327 – Database Design and Modeling

Final Project

04/21/2023

Team 1

Jaden Dinh, Kyle Duong, Robert Dang, Samson Mulugeta, Alice Sun

Introduction

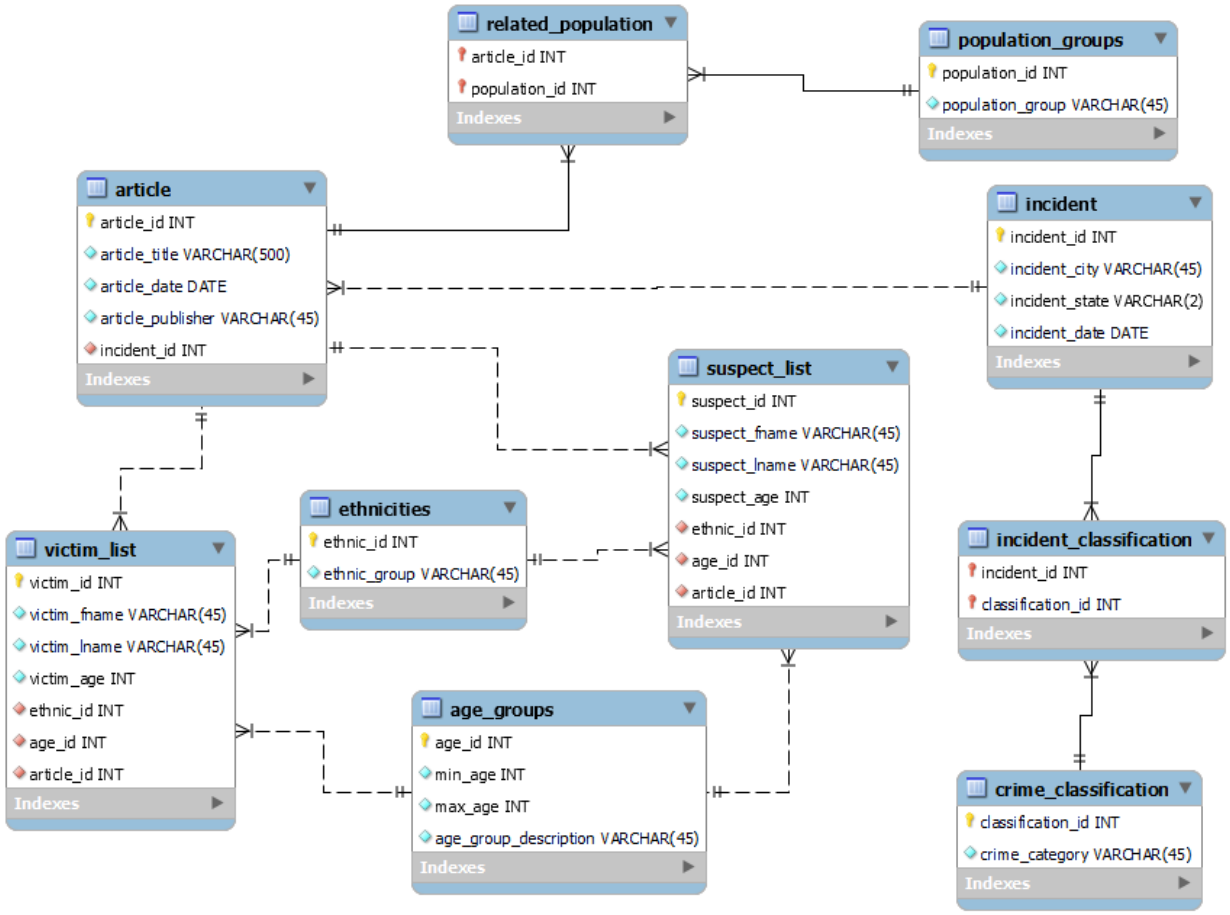
Hate crimes have always been a prevalent issue throughout history and that holds true even today. Of course, despite the times, hate crime all constitutes the same characteristics and the severity of it can range from anything minor like verbal harassment to physical abuse. Recently, we have seen a rise in hate crimes against people of color, especially within the Asian and black communities. Since our group's demographic consists of those two groups, we want to take this as an opportunity to dive deeper into the topic and use our experiences with hate crimes to develop a meaningful database.

As a group, we want to develop a database that would provide insight into the severity and prevalence of hate crimes within today's society. This database will also contain the different forms that hate crimes could take shape and manifest. Another objective for this database is to bring forth the urgency of the issue and how we need to start taking action to protect people from future incidents like the ones noted within the database. To accomplish this, we will use the data available from the dataset provided, which pulled information from ProPublica. It will hold a wide variety of hate crimes reported by the public, combined with information from law enforcement, news sources, social media, and non-profit organizations. It will also provide other attributes like the date of the report, general location, and some key information.

For the purpose of this project as well as to make the database fit into our objective, we will look into modifying some of the attributes to make the tables easier to understand. Specifically, we might consider separating some attributes like publications and locations into their own separate table. Most of the attributes in the dataset are useful, but some would benefit from having their own table to give the viewer a better understanding of what aspect of our database they're viewing. Also, there are some columns within the original dataset that do not contain any row of data, which we will take out or turn into something useful and related to our database.

Database Description

Logical Design:



Physical Database:

Our organization seeks to create a comprehensive database that illuminates the scope and gravity of hate crimes in contemporary society, with a special emphasis on crimes committed against those who are within the BIPOC communities. We want to provide insight into the severity and prevalence of hate crimes against BIPOC within today's society. This database will also contain the different classification that would constitute a hate crime. Another objective for this database is to bring forth the urgency of the issue and how we need to start taking action to protect people from future incidents like the ones noted within the database. The database will include a number of tables, including ones for the articles, suspects and victims involved, the incident, population groups, and some other aspects. Information about each occurrence, including the date, the kind of hate crime, and the location will be listed in the table of hate crime incident. The article table will provide details on the news media and nonprofits that covered the tragedy. On top of that, there will also be separate tables to document suspects and victims involved, including descriptors like their names, ethnicities, ages, etc. Our data will seek to fill

the gap within hate crime reporting performed by the FBI, and attempt to highlight crimes that might have gone unreported.

Sample Data:

We will utilize the dataset provided by ProPublica, which encompasses comprehensive information on hate crime reports from various sources, including law enforcement agencies, journalistic organizations, social media platforms, nonprofits, and reports submitted by the general public. To ensure the effectiveness of our database, we will make necessary modifications to the ProPublica dataset and organize it into tables based on relevant attributes. Additionally, we will thoroughly clean the dataset by eliminating rows with insufficient data, particularly those lacking information in certain columns. Furthermore, we have excluded articles that primarily focus on general topics rather than specific hate crime incidents, as they do not align with the primary objective of our database.

Views/Queries:

Query Name	**Join (4)**	**Filter (3)**	**Aggregate (2)**	**Linking Table (1)**	**Subquery (1)**
crime_classification_analysis	X		X		
murder_victims_analysis	X	X	X	X	X
april_suspects	X	X			
young_adult_suspects	X	X	X		X
pct_of_suspects_by_ethnicity	X		X		

The following explains what each of our queries displays:

- Query 1: Create a view that shows the total number of incidents within each crime category.
- Query 2: Create a view that shows the number of victims for murder incidents and their average age.

- Query 3: Create a view that lists all the suspects involved in April incidents along with their age.
- Query 4: Create a view that details the number of suspects that are young adults among the total of suspects within each articles.
- Query 5: Create a view that shows the percentage of suspects by ethnicity

Changes From the Original Design

We have discovered certain adjustments that will improve the database's efficiency and usability after thoroughly examining the database's initial architecture. The improvement of attribute separation is one significant modification we suggest. As was stated in the initial design, breaking off attributes like article, suspects, and victims into their own tables can help make the database easier to grasp. We can better organize the data and make it possible for research based on particular elements of hate crimes by separating the incident and crime_category into distinct tables, then linking them via a child table. This change will make the database's overall organization and usability better. We also added numerous table that are used for classification/categorization like ethnicity and age_group. This will help a more efficient view of the information and better connect different aspects together. Besides those changes to the database, we also restructured some of our table to better store our data for our purposes. We also revert our changes to create linking table for between article and each of the suspect and victim tables. We realized that with our aim to only report unique incidents, combine with the small scope of our database, there will not be repeating articles for the same suspects or victims. Thus, this reduce the relationship between article and suspects or article and victims to a one-to-many relationship, which eliminates the need for a linking table. Last but not least, with regard to data sources, we will keep using the dataset given by ProPublica, which contains data from a variety of trustworthy sources, including law enforcement, journalism organizations, social media, and nonprofits. To uphold openness and intellectual property rights, we will make sure that the data sources are properly cited and attributed. By making these adjustments, we hope to produce a more complete and user-friendly database that clearly illustrates the prevalence and seriousness of hate crimes, offers insightful information, and inspires action to address this critical societal problem.

Database Ethics Considerations

It's critical to think about and solve any potential legal and ethical issues that might surface as we work on our database project. This entails ensuring that data is gathered and used lawfully and morally, safeguarding the privacy and confidentiality of persons, abiding by fair use and copyright rules, and taking into account any potential effects on the communities and people the database relates to. We can produce a useful, responsible, and ethical resource by approaching the endeavor with consideration and ethics. We will be gathering is public and comes from the non-profit organization, ProPublica. This means that if our database were to be breached, there would be little to no data privacy concerns because all the information was already accessible online. While the data may be legally obtainable, it is still important to consider data privacy and ensure that sensitive information is not misused, especially in a topic such as documenting hate crimes. For this reason, we will try to be as impartial as possible and use objective language in our project. As of now, we do not plan on acquiring data from any proprietary sources. We will cite all our information from sources and databases that we use on the web.

Lessons Learned

There have been some important lessons discovered while creating and using the hate crime database. First and foremost, we understand the important of a proper managed database as well as carefully review of queries before it execution. If the relationships between tables are not defined correctly, it could lead to conflict or misrepresentation of data and will throw the table off tremendously, much like what happened during our initial data insert process. Also, before making changes to data, table, or even the database itself, proper review is important to avoid mistakes that could put the database process into jeopardy. On a bigger scope, we now understand how crucial diversity and collaboration are. We can collect a greater diversity of opinions and experiences by incorporating people and organizations from many backgrounds, including those who are directly impacted by hate crimes. This cooperation guarantees that the database is thorough, accurate, and considerate of the communities it seeks to serve. Additionally, it promotes a sense of empowerment and ownership among those participating, encouraging active engagement in combating the problem of hate crimes. The importance of data integrity and quality assurance is a further lesson that can be drawn. Verifying and

cross-referencing the data from many sources is essential to guarantee its accuracy and dependability. This procedure might include fact-checking, spotting any biases or inconsistencies, and putting validation measures in place. For the database to remain current and relevant, regular updates and maintenance are also required. Ethics and privacy concerns have come to light as a major project lesson. We are aware of the sensitivity of the data on hate crimes and the possible harm that could result from its improper use. The data must therefore be handled with the highest respect for confidentiality, privacy, and ethical considerations. A messed up within the database could lead to false information and potentially lead to bigger consequences in real life. This project also gave us an opportunity to strengthen our normalization skills using a database that we created ourselves. The result turned out to be a success and we were able to improve on our understan of the difference between each normalization steps.

Potential Future Work

Potential future work that could be done with this is expanding our hate crime database to more than just the United States. We could spread throughout different locations around the world, which could help other countries that deal with hate crimes. In order to do that we could partner up with the local law enforcement agencies, human rights groups, and other relevant stakeholders in the countries to ensure accurate and comprehensive data collection. We could also enhance access to it by translating the database and its interfaces into multiple languages. This would allow individuals and organizations from different regions to easily access and contribute to the database. It also helps people who speak other languages to understand the database. Another thing we could work to implement is a real-time, or at least close to real-time, system that would update the database as reports are made each day.