Khaleel Denton
Magdaly Abrego
Girija Mallampalli
Nimbus Kunwar
Final Project Database + Report
INST327 - 0101 - Group 1
12/09/2024

Introduction

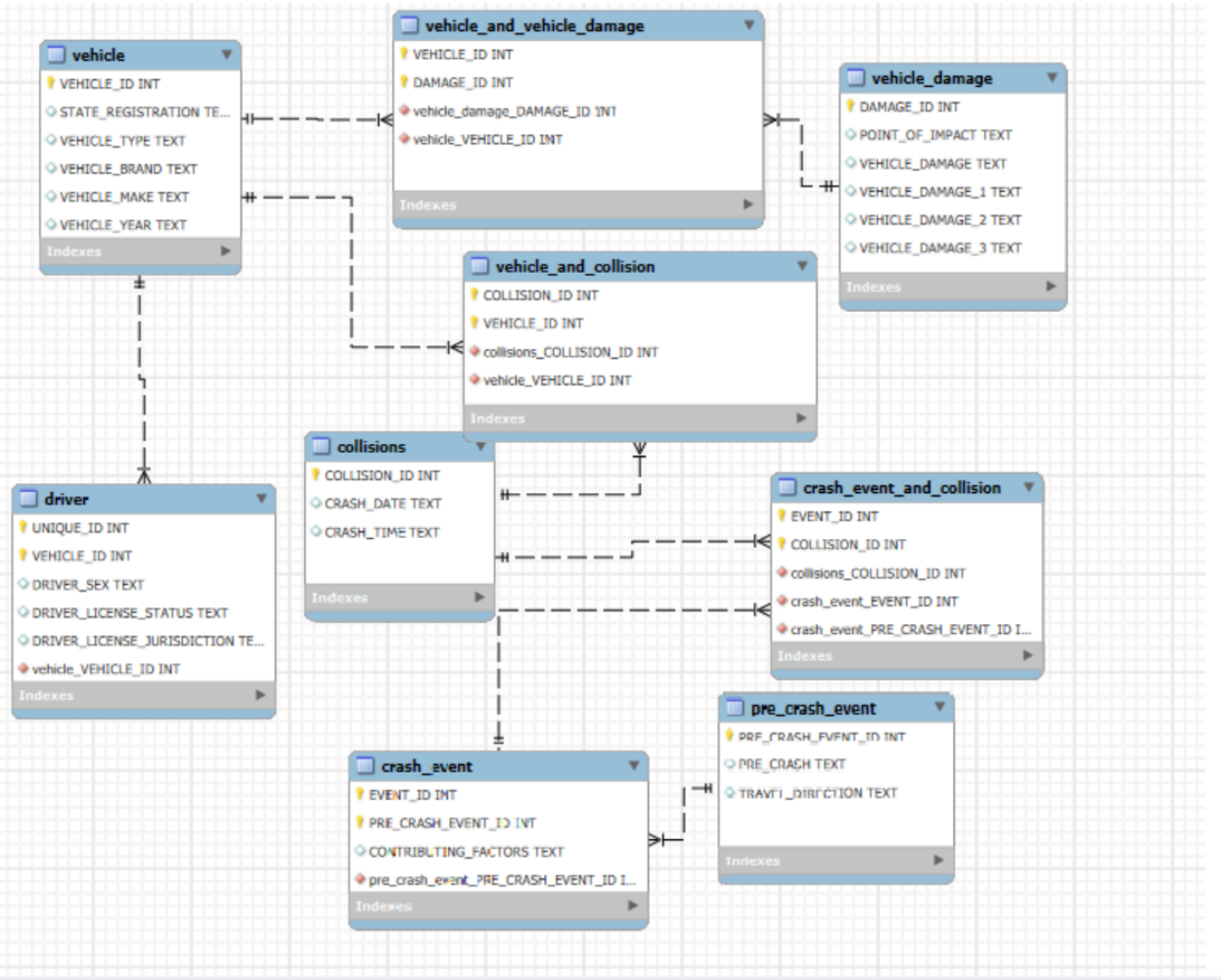
The Motor Vehicle database, provided by the New York Police Department, is designed to give a comprehensive summary and overview of traffic incidents involving cars during a certain time period. It provides a detailed record of events involving vehicles on the road. This database contains various data elements, such as Unique ID, collision ID, crash date, crash time, and information about the vehicles involved, including make, model, year, and type. It also tracks crucial driver information, such as driver gender, license status, and jurisdiction, as well as crash specifics, such as contributing variables and locations of impact. By accumulating this data, the database aims to provide useful information on the incidence and causes of motor vehicle collisions during a specific time frame.

The project scope is to enhance roadway safety by analyzing car crashes over a period of time. This analysis focuses on understanding how different factors such as driver demographics, vehicle characteristics, and other external conditions contribute to motor vehicle accidents. Addressing these issues is very important, as traffic collisions have an impact on public safety. They result in loss of life, injury, and high economic costs. By using the data, the project seeks to inform policy decisions, enhance driver education, and in the end, hopefully, ensure safer driving behaviors.

The Motor Vehicle Collisions database is mainly applied by researchers, lawmakers, and safety groups to look into the causes and effects of vehicle crashes. This data information helps us evaluate road safety concerns and make final, educated decisions. In addition, insurance providers such as Geico, State farm, Progressive and other organizations can potentially use the data to evaluate risks and create safer vehicles. The database helps lower collisions and boost safety on the road.

Database Description

- **Logical Design**



- **Sample Data:**

We have a total of 9 tables within our database. A major table within the database is the ‘Collisions’ table. Collision_ID being the primary key, and CRASH_DATE and CRASH_TIME being the attributes.

	COLLISION_ID	CRASH_DATE	CRASH_TIME
	100201	9/7/12	9:03
	4213082	9/23/19	8:15
	3307608	10/2/15	17:18
	3308693	10/4/15	20:34
	297666	4/25/13	21:15
	3434155	5/2/16	17:35
	4229067	10/24/19	13:15
	3503027	8/18/16	12:39
	196425	7/16/13	11:20
	2975897	11/26/12	18:12
	3487936	7/22/16	15:40
	3268858	8/1/15	8:17
	3499697	8/13/16	21:05
	4229563	10/21/19	17:55
	4322767	6/6/20	18:30
	176016	8/11/12	19:36
	3516125	9/8/16	14:40
	79561	4/9/13	15:10
	3412270	4/1/16	15:46
	4229538	10/24/19	17:30
	4228839	10/24/19	16:00
	19615	7/16/13	17:10

- **Views/Queries:**

View Name	Req. A	Req. B	Req. C	Req. D	Req. E
collisions_and_vehicles_view	X	X		X	
collision_count_by_vehicle_type	X	X	X	X	
drivers_in_collisions	X	X			

frequent_time_of_crash	X		X	X	
vehicle_damage_details	X	X		X	X

Changes From Original Design:

Since the initial proposal, there have been several significant changes to our project. One of the key changes involves the number of tables and their structure. Initially, we planned to use fewer tables, believing that additional ones weren't necessary. However, after receiving feedback from our TAs, we recognized that including more tables would help create a more organized and concise dataset, as it was indeed required. Through multiple iterations of feedback and revisions, we successfully developed our ERD. In terms of table attributes, we have removed any redundant entities, which facilitated the normalization of the tables, as shown in our ERD.

As we delve further into this project, we are beginning to appreciate how such a database can be utilized for various data analysis and decision-making purposes. For example, we can answer questions related to vehicles, such as, "What types of vehicles are most frequently involved in collisions?" Additionally, we can explore more specific inquiries, such as, "What are the peak hours or days for collisions?" The database also allows us to analyze trends related to specific regions and demographics, enabling us to report these findings in an ethical and responsible manner.

Database Ethics Considerations:

Not much has changed regarding the data, privacy, and ethical use considerations since the start of this project. Our team continues to remain vigilant in ensuring the ethical and

responsible management of the NYPD traffic incident database. This data involves a sensitive topic—both emotionally, as it may impact individuals personally affected by such incidents, and in terms of privacy and security concerns.

The database in use includes brief details about when and how these incidents occurred, specifics about the vehicles involved, and information such as driver's license details and the impacts of incidents. It does not, however, include any sensitive personal identifiers such as the names of victims or those involved in the incidents. As the database is openly available for public access and does not require additional proprietary or restricted data, it aligns with ethical and legal considerations for its use in our project.

Even though the dataset lacks directly identifiable information, privacy and security remain top priorities. If any sensitive elements, such as personal names, were included, they would not be stored in our analysis environment. The provided data is sufficient to learn from, analyze, and use effectively for this semester-long group project while upholding standards of copyright, fair use, and ethical responsibility.

Lessons Learned:

The Entity-Relationship Diagram (ERD) was a large-scale project and according to the group, it created confusion and inconsistencies in database design. The key challenge was mapping the relationships into a normalized diagram that accurately represented the database structure. In response to this, the team consulted with multiple TAs, designed the project through a step-by-step cycle, and preprocessed the MySQL database to reduce complexity while maintaining data quality. This approach ensured consistency with project goals.

Another lesson we learned throughout the process of this project was how to properly design our ERD. Initially, we believed that reducing the number of tables would simplify the connections between them and make the database more efficient. However, we quickly realized that this approach often led to data redundancy and inconsistencies, complicating queries and analysis. This structure not only improved the clarity of the database but also aligned better with the principles of database design. This adjustment taught us the value of prioritizing clarity and normalization over simplicity when designing complex systems.

Potential Future Work:

With more time, our database can be expanded to include more tables and data. This expansion can offer a more comprehensive understanding of collisions and their contributing factors. For example, incorporating new tables such as weather conditions, road conditions, and geospatial data, the database could aid in detailed analyses of how environmental and location-based factors influence accident rates. An expanded database would also help users answer more complex questions. It could evaluate the effectiveness of road safety policies, determine which vehicle features (e.g., safety ratings or fuel types) correlate with lower accident rates, and analyze driver behavior trends over time. These capabilities would make the database invaluable for city planners aiming to improve infrastructure, insurance companies assessing risk, and law enforcement agencies investigating traffic trends.

Some future improvements to the database would involve creating additional tables and enhancing linking tables to allow for more seamless connections between tables and data. By adding new tables, the database can capture a wider range of information, providing a more

comprehensive and interconnected view of the relationships between collisions, vehicles, drivers, and other factors.

References:

(NYPD), P. D. (2024, September 27). *Motor vehicle collisions - crashes: NYC open data*. Motor Vehicle Collisions - Crashes | NYC Open Data. https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/about_data