Final Report: Montgomery County Motor Vehicle Theft Analysis Database (2020–2025)

Group 304
Tylor Davis, Dylan Yoon, John Brusie, Kenia Aviles, Jonathan Bui

---

1. Introduction

Theft-related crimes, particularly those involving motor vehicles, have been a persistent issue in Montgomery County over the past several years. These types of crimes not only disrupt the lives of the individuals affected but also strain community trust and law enforcement resources. In response to these concerns, this project set out to design and implement a relational database that could store, manage, and analyze motor vehicle theft data from January 1, 2020, to February 23, 2025. The overarching goal of this initiative is to create a system that can serve multiple stakeholders by offering a centralized and queryable platform to examine patterns and trends in vehicle-related crime.

A key motivation for focusing on motor vehicle theft is its frequency and societal impact. Stolen vehicles often support other criminal activity, and areas with high rates of motor vehicle crime may experience additional public safety concerns. By analyzing offense codes, crime classifications, locations, and timestamps, the database aims to empower data-driven decisions for crime prevention, smarter allocation of policing resources, and well-informed policy development. In addition to providing statistical insights for institutional users, the database is also intended to foster community awareness, engagement, and ultimately a shared responsibility for public safety.

The Montgomery County Motor Vehicle Theft Database bridges a critical information gap by presenting curated, up-to-date information from multiple sources in a coherent schema that reflects the complexity of real-world criminal activity. It not only acts as a digital record but also as a tool for cross-disciplinary collaboration between law enforcement, public officials, academic researchers, insurance agencies, and residents. This report outlines the technical architecture, implementation process, challenges encountered, and future opportunities for improving and expanding the database system.

---

2. Database Description

2.1 Logical Design

The logical design structures the essential elements of theft-related crimes in Montgomery County into a set of normalized, relational tables. These tables include information about each crime, crime types, offenses, locations, cities, places, police districts, and dispatch information.

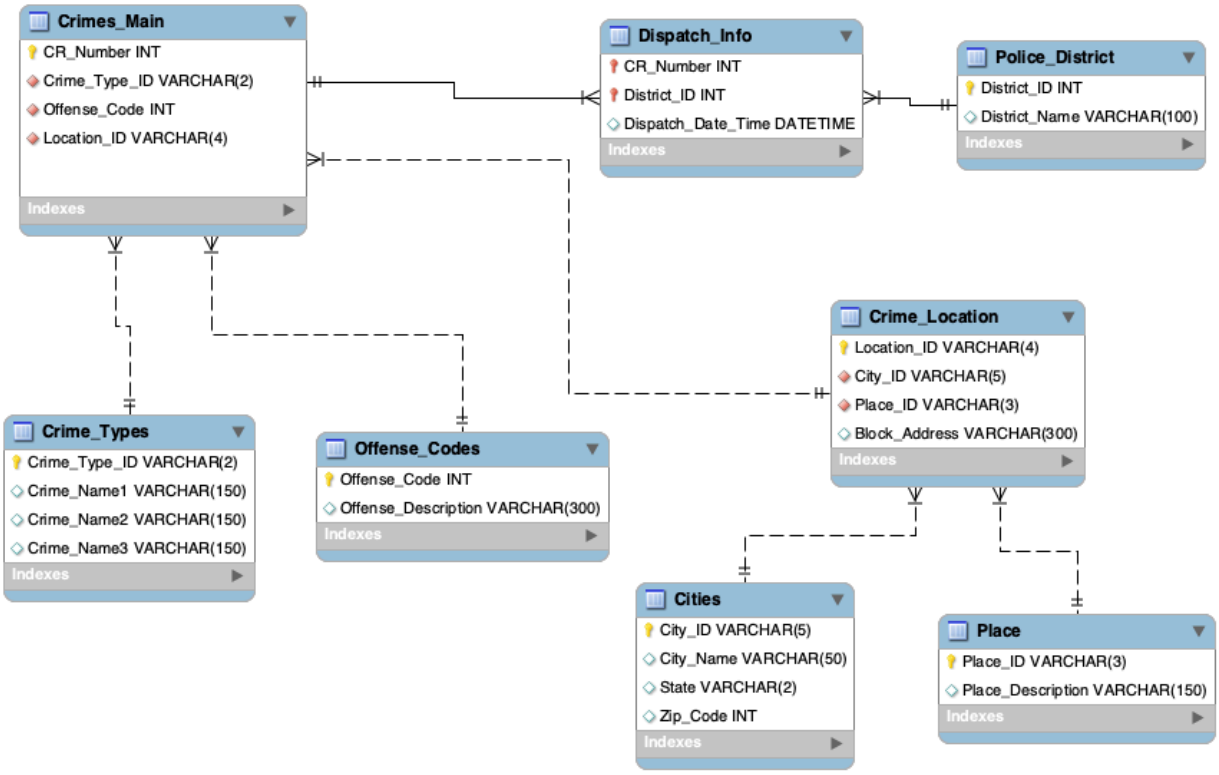
The Crimes_Main table serves as the foundation of the database; it records each theft case with a unique case number (CR_Number) and establishes relationships with several supporting tables through foreign keys.

- Crime_Type_ID connects to the Crime_Types table, which categorizes crimes using multiple descriptive levels (Crime_Name1, Crime_Name2, Crime_Name3).
- Offense_Code links to the Offense_Codes table, providing more context to the type of crime committed.
- Location_ID connects to the Crime_Location table, which further breaks down into City_ID and Place_ID, referencing the Cities and Place tables. This setup provides a geographic context of the crime.

The Dispatch_Info table tracks the police response time detail for each crime, including District_ID and the exact Dispatch_Date_Time. The District_ID is a foreign key to the Police_District table, which holds the names of the districts, which will show how different districts are responding to theft incidents.

All the tables are connected using foreign keys and use standardized formats. The design follows normalization rules to avoid duplicate data and keep data organized. It makes it easier to run

complex queries, and overall, the structure supports in-depth analysis of theft trends.



2.2 Physical Database

The database was built in MySQL Workbench. Tables were created according to the final ERD schema, using appropriate data types and constraints. Foreign keys enforce relationships between tables, such as linking CrimeEvent to Place and PoliceDistrict. All tables were populated with sample data, and the entire schema was backed up into a single .sql file to ensure portability. The backup was successfully tested by restoring it on a second system.

2.3 Sample Data

Each primary (non-join) table contains at least 30 rows of data, while the join tables do hold considerably less. Due to our scope being narrowed to focus on auto theft in one specific city, some of the details are limited. Tables such as police district and crime types are limited to only 1 or a dozen options since our crime data is niche. Our data does house essential details for each case, including the time of dispatch of officers, the address, the place of theft, and accurate crime descriptions.

For example:

- Offense Code 2404 corresponds to "Motor Vehicle Theft - Vehicle Theft."
- Offense Code 2305 corresponds to "Theft From Motor Vehicle."

Place descriptions were categorized (e.g., "Parking Garage," "Retail," "Residence") and linked to records to support queryable spatial and contextual analysis. Timestamps span the designated five-year range to reveal seasonal or time-of-day patterns.

2.4 Views/Queries

The database includes the following five views, each demonstrating important SQL concepts:

- **View 1: Crime stats by City and State**
    - JOIN (Crime_Location + Cities)
    - FILTER NOT NULLS
    - AGGREGATION (COUNT, GROUP BY, ROUND)
    - **Meets Requirements A, B, C**
- **View 2: Crimes in Zip 20866**
    - JOIN (Crime_Location + Cities)
    - FILTER (WHERE Zip_Code = '20866')
    - **Meets Requirements A, B**
- **View 3: Streets with Multiple Crimes**
    - JOIN (Crime_Location + Cities)
    - FILTER WHERE Block_Address IS NOT NULL/empty
    - AGGREGATION (COUNT + HAVING)
    - **Meets Requirements A, B, C**

- **View 4: Location and City Details**
    - JOIN (Crime_Location + place + Cities)
    - No filtering or aggregation
    - Includes linking table-style structure Crime_Location linked to both Place and City
    - **Meets Requirements A, D**

- **View 5: MD Crime Locations**
    - JOIN (Crime_Location + Cities)
    - FILTER (State = 'MD')
    - **Meets Requirement A, B**
- **View 6: Subequery - High Crime Streets in Top Zip Codes**
    - JOINs between Crime_Location and Cities
    - Filter via WHERE and HAVING
    - Aggregation with COUNT and GROUP BY
    - Uses a subquery (with a nested SELECT)
    - **Meets Requirements, A, B, C, E**

View Name	A	B	C	D	E
Crime stats by City and State	☑	☑	☑	☐	☐
Crimes in Zip 20866	☑	☑	☐	☐	☐
Streets with Multiple Crimes	☑	☑	☑	☐	☐
Location and City Details	☑	☐	☐	☑	☐
MD Crime Locations	☑	☑	☐	☐	☐
High Crime Streets in Top Zip Codes	☑	☑	☑	☐	☑

---

3. Changes from Original Design

Originally, the database proposal considered a broader category of theft-related crimes. However, during implementation, the scope was narrowed to focus exclusively on motor vehicle theft crimes in the Burtonsville area and crimes that occurred in places considered ‘residential.’ We decided to focus on Burtonsville, as although it only accounted for ~20% of the original data, Burtonsville has a large residential area and has experienced all forms of auto theft crime. We had to make a few changes to our ERD and switch around columns in certain tables to fix

redundancies and maintain only relevant information. In addition, we had to fix a connection issue between the Crimes Main table and the Crimes Location table.

---

4. Database Ethics Considerations

This project was designed with careful consideration of database ethics, especially regarding bias and data privacy. Since the dataset originates from police reports, potential over-policing in certain communities may skew results. To mitigate this, the project intentionally avoids demographic data like race or age. No personally identifiable information is stored.

Efforts were made to balance representation across locations and crime categories. Place descriptions are generalized (e.g., parking lot, residence) to maintain anonymity. Moreover, only publicly available data was used, and it was cited properly to ensure ethical compliance.

---

5. Lessons Learned

Key challenges included importing and cleaning the dataset, handling inconsistencies in place and crime names, and writing advanced queries using JOINs and subqueries. Developing meaningful views required iterating on both query logic and normalization. Through troubleshooting and peer feedback, we improved the structure and readability of the schema.

Working in MySQL Workbench also presented technical hurdles, particularly around foreign key constraints and exporting reliable backups. However, these challenges led to a deeper understanding of relational integrity and the value of rigorous testing.

---

6. Potential Future Work

Future development could include:

- Integrating visualization tools like Tableau for dashboards
- Expanding data scope to include demographic or socioeconomic overlays
- Connecting to a live data feed for real-time monitoring
- Creating user roles and access control for different stakeholders

These enhancements would make the database even more impactful for analysis, planning, and public awareness.

---

7. Citations

Montgomery County Police Department. (n.d.). *Crime data for Montgomery County* [Data set]. Montgomery County Government. Retrieved [Month Day, Year] , from https://data.montgomerycountymd.gov/Public-Safety/Crime/icn6-v9z3/data_preview