

## **Final Project Written Report**

### **Introduction:**

Our project revolves around the Montgomery County Crime Database, a comprehensive repository of crime-related data in Montgomery County, Maryland, which is the most populous county in the state. With a population exceeding one million residents and being part of the Washington D.C. metropolitan area, Montgomery County is notable for its affluence, high average household income, and a significant percentage of residents holding post-graduate degrees. In regards to the Montgomery County Police Department (MCPD), they are responsible for maintaining law and order in this thriving community, comprising approximately 1,300 officers and 650 support personnel.

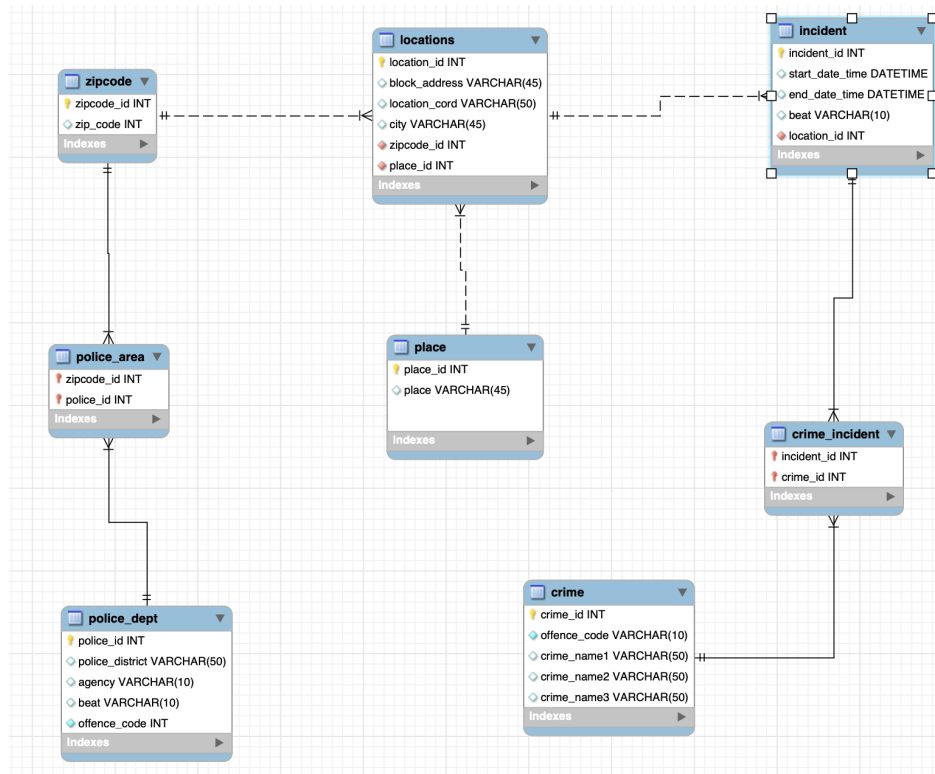
The Montgomery County Crime database includes a significant amount of information on open cases, reported and attempted crimes, and closed cases, starting from July 16, 2016, onwards. This dataset is updated daily and is managed by the EJustice system, which is the records-management system used by MCPD. With over 300,000 cases and daily updates, it offers a snapshot of criminal activity in the county. Our project's scope involves using this dataset to gain insights into crime trends, patterns, and demographics in Montgomery County.

We chose this topic because as residents of Montgomery County and adjacent areas, it addresses a relevant aspect of community well-being and public safety. Also, analyzing crime data from such a populated and affluent county can help to identify areas of concern, allocate resources efficiently, and implement strategies to reduce crime rates. It can also serve as a valuable resource for law enforcement agencies, policymakers, researchers, and the general public interested in understanding and addressing crime in their community.

### **Database Description:**

For our project, utilized a database containing Montgomery County crime data from 2022. This database included publicly available crime data which can also be found on the *dataMontgomery* county website. The target audience of this database, primarily Montgomery

County inhabitants and workers, may benefit from this database as our mococrime backup database pinpoints specific details about crimes by type, setting, city, recurrence etc. We intended to answer questions like: “*which settings will likely face recurrences of certain crimes*”, or “*which police districts reported to the most number of crimes*” We also revised our ERD based on feedback received on the Project Logical Design. Attached is our mocorime ERD:



### ***Physical Database***

Our database is named mocorimes, and includes all sample data, tables, and views; found in our backup database. Here are examples of queries we constructed to create the views shown below in “Views”:

```
/* Query 6: Crimes that Beats Responded To */
```

```
• DROP VIEW IF EXISTS mococrime.BeatResponseToCrimes;
• CREATE VIEW mococrime.BeatResponseToCrimes AS
SELECT z.zip_code AS zipcode,
       l.city,
       p.beat AS beat,
       c.crime_name2 AS crime,
       c.crime_name3 AS details,
       COUNT(c.crime_id) AS total_cases
FROM   mococrime.police_dept p
       JOIN mococrime.crime c ON p.offence_code = c.offence_code
       JOIN mococrime.incident i ON p.beat = i.beat
       JOIN mococrime.locations l ON i.location_id = l.location_id
       JOIN mococrime.zipcode z ON l.zipcode_id = z.zipcode_id
WHERE  p.beat IS NOT NULL
GROUP BY z.zip_code, l.city, p.beat, c.crime_name2, c.crime_name3
ORDER BY total_cases DESC, city;
```

```
/* Query 3: Setting types with more crimes compared to others */
```

```
• DROP VIEW IF EXISTS mococrime.HighCrimeSettings;
• CREATE VIEW mococrime.HighCrimeSettings AS
SELECT p.place,
       COUNT(ci.crime_id) AS case_count
FROM   mococrime.crime_incident ci
       JOIN mococrime.crime c ON ci.crime_id = c.crime_id
       JOIN mococrime.incident i ON ci.incident_id = i.incident_id
       JOIN mococrime.locations l ON i.location_id = l.location_id
       JOIN mococrime.place p ON l.place_id = p.place_id
WHERE  p.place IS NOT NULL
GROUP BY p.place
ORDER BY case_count DESC;
```

```
/* Query 7: Crimes generally committed in Montgomery County */
```

```
DROP VIEW IF EXISTS mococrime.CrimesInMontgomery2022;
CREATE VIEW mococrime.CrimesInMontgomery2022 AS
SELECT YEAR(i.start_date_time) AS year,
       c.crime_name2 AS crime,
       COUNT(c.crime_id) AS total_cases
FROM   mococrime.incident i
       JOIN mococrime.crime_incident ci ON i.incident_id = ci.incident_id
       JOIN mococrime.crime c ON ci.crime_id = c.crime_id
WHERE  YEAR(i.start_date_time) = 2022
GROUP BY YEAR(i.start_date_time), c.crime_name2
ORDER BY crime;

SELECT * FROM mococrime.CrimesInMontgomery2022;

DROP VIEW IF EXISTS mococrime.CrimesComparison;
```

### *Sample Data*

For this project, we utilized the Montgomery County Crime.csv file to analyze data. This dataset consisted of reported crimes and open/closed cases in Montgomery County, Maryland. The crimes and cases in this database are dated back to July 16, 2016. As the original dataset is rather large with 327,667 rows, we decided to define the scope of our analysis to the past year to ensure efficiency and relevance. Specifically, we narrowed our database to focus on records from January 2022 through December 2022, so that we may compare and infer insights between last year. In limiting our timeframe to crimes reported in 2022, our database is better concentrated on reflecting recent crimes and their potential recurrence.

In addition to the temporal filter, we did not use all the columns originally provided and instead selected the most pertinent attributes for our database to emphasize. Attributes like crime name, location, time, and incident id were used in our database, as we excluded less relevant attributes like longitude, latitude, and street prefix. Lastly, we created seven tables representing

the logical design of our model where all tables were filtered to show only the first 500 rows of information for us to then create our database.

Using these data reduction techniques and class instruction requirements, we accomplished our goal of creating a refined and thorough database that aligned with our objectives. On a larger scale, this dataset could be facilitated for improved efficiency and analysis, allowing legislators and residents to infer meaningful insights into our created queries about crime in Montgomery County. Here are the tables we created in SQL:

- **Table 1: Zipcode**

- Entities**

- Zipcode\_id (PK)
    - Zip\_code

- **Table 2: Locations**

- Entities**

- Location\_id (PK)
    - Block\_addresses
    - Location\_cord
    - City
    - Zipcode\_id (FK)
    - Place\_id (FK)

- **Table 3: Incident:**

- Entities**

- Incident\_id (PK)
    - Start\_date\_time
    - End\_date\_time
    - Beat
    - Location\_id (FK)

- **Table 4: Police\_dept**

- Entities**

- Police\_id (PK)
    - Police\_district
    - Agency

- Beat
- Offence\_code (CPK)
- **Table 5: Place**
  - Entities**
    - Place\_id (PK)
    - Place (CPK)
- **Table 6: Police\_area**
  - Entities**
    - Zipcode\_id (PK)
    - police\_id (CPK)
- **Table 7: Crime**
  - Entities**
    - Crime\_id (PK)
    - Offence\_code (CPK)
    - Crime\_name1
    - Crime\_name2
    - Crime\_name3
- **Table 8: Crime\_incident**
  - Entities**
    - Incident\_id (PK)
    - Crime\_id (FK)

### *Views from Queries:*

We had six views in our backup database. Here is the requirements table, as well as a few examples of views in our mococrime database.

View Name	Req A	Req B	Req C	Req D	Req E
Query 1 - CitiesMostCrime	X	X	X		
Query 2 - CitiesLeastCrime	X	X	X	X	X
Query 3- HighCrimeSettings	X		X	X	
Query 4- HighCrimeCities	X		X	X	X
Query 5- MostCrimePoliceDistrict	X	X	X	X	X
Query 6- BeatResponseToCrimes	X			X	
Query 7- CrimesinMoco2022	X	X	X		

zipcode	city	beat	crime	details	total_cases
20850	ROCKVILLE	1A3	All Other Offenses	POLICE INFORMATION	32
20902	SILVER SPRING	4L2	Shoplifting	LARCENY - SHOPLIFTING	15
20902	SILVER SPRING	4L2	All other Larceny	LARCENY (DESCRIBE OFFENSE)	15
20874	GERMANTOWN	5N1	All Other Offenses	POLICE INFORMATION	12
20874	GERMANTOWN	5N1	Driving Under the Influence	DRIVING UNDER THE INFLUENCE LIQUOR	12
20886	MONTGOMERY VILLAGE	6R1	Credit Card/Automatic Teller Machine Fraud	FRAUD - ILLEGAL USE CREDIT CARDS	12
20850	ROCKVILLE	1A3	All Other Offenses	LOST PROPERTY	12
20902	SILVER SPRING	4L1	Simple Assault	ASSAULT - 2ND DEGREE	12
20902	SILVER SPRING	4L2	Destruction/Damage/Vandalism of Property	DAMAGE PROPERTY (DESCRIBE OFFENSE)	10
20877	GAITHERSBURG	6P3	Driving Under the Influence	DRIVING UNDER THE INFLUENCE LIQUOR	8
20874	GERMANTOWN	5N2	Driving Under the Influence	DRIVING UNDER THE INFLUENCE LIQUOR	8
20850	ROCKVILLE	1A3	Destruction/Damage/Vandalism of Property	DAMAGE PROPERTY - PRIVATE	8
20876	GERMANTOWN	5M1	All Other Offenses	LOST PROPERTY	6
20854	POTOMAC	1B1	All Other Offenses	SUDDEN DEATH	6

place	case_count
Street - In vehicle	7
Residence - Apartment/Condo	6
Residence - Single Family	5
Other/Unknown	5
Residence - Nursing Home	3
Gas Station	2
Parking Lot - Residential	2
Retail - Department/Discount Store	2
Auto Repair	1
Retail - Drug Store/Pharmacy	1
School - Elementary/Secondary	1
Retail - Other	1
Parking Lot - Commercial	1

year	crime	total_cases
2022	Aggravated Assault	1
2022	All other Larceny	3
2022	All Other Offenses	13
2022	Burglary/Breaking and Entering	3
2022	Counterfeiting/Forgery	1
2022	Credit Card/Automatic Teller Machine Fraud	2
2022	Destruction/Damage/Vandalism of Property	5
2022	Driving Under the Influence	4
2022	Drug/Narcotic Violations	2
2022	False Pretenses/Swindle/Confidence Game	1
2022	Robbery	1
2022	Runaway	1
2022	Shoplifting	3
2022	Simple Assault	4
2022	Theft from Building	1
2022	Theft From Motor Vehicle	1

### **Changes from the original design:**

Since our initial proposal and progress report submissions, we have made notable changes to the entities in our logical design. The feedback we received noted our absence of setting composite keys in our linking tables and some columns as not null. Through discourse with one another and the instructional team, we have changed our attributes (tables) and keys. We have now identified the following 8 attributes: Zipcode, Locations, Incident, Police\_dept, Place, Police\_area, Crime, and Crime\_incident. We decided to eliminate the Address table and most of its keys and instead created a Zipcode and Place attribute. These changes reflect database design best practices and improve the relationships between attributes for implementation into SQL. In addition to design changes, we've also been able to successfully create our physical database on the MySQL server, input sample data, and create sample queries for views.

### **Database Ethics Considerations:**

In developing our proposed database to assist residents and legislators of Montgomery County, we recognize the importance of ensuring inclusivity and upholding ethical considerations. While our database can serve as a helpful tool to residents and legislators of Montgomery County, it is crucial to ensure inclusivity by capturing and retaining the full social, historical, and demographic diversity of the area. The given database already includes variables such as crime types, locations, dates, and times. Despite this data being available via public records, the team has taken into account the consideration fair use and ethical use of the database we're designing. To maintain ethical standards, we have taken measures to safeguard the exact names and addresses of the locations where crimes took place. Additionally, we continue to promote ethical use of the data by transparently providing the database sources, outlining any historical and social contexts that should be taken into consideration when analyzing the data, and publicly outlining the intended use of the database to ensure it is responsibly and fairly used.

### **Lessons Learned:**

From encountering this project as a group, it was understood that this was our first experience learning about SQL queries and how to design them to best communicate what our data was representing. Our first moment, where we became “stuck” was during the Top-Down Logical design deliverable. We realized that waiting til the day of the deadline was not a realistic or efficient method, and since then we have fixed the issue by having our iMessage group chat as a haven for communication as well as holding regular meetings to update each other on project progress marks for contributions to upcoming deliverables. Our issue with the specific deliverable was being able to best organize what parts of the database and how to manipulate these different aspects to be able to answer our question. With a full day of discussion, we got that done. Now with the following deliverables such as the Project Logical Design and the Progress Report, we had the timing, communication, and organization down. However we were having issues with normalization and creating our ERD tables, and we soon realized that to perform optimally, we would need help from the instructional team. This resource improved our skills related to normalization and finding relationships between attributes thus improving our understanding of database design. All in all, we learned about the value of leveraging available resources and knowing when to seek assistance from the instructional team which had a positive impact on the progress of the project. On a final note, as a team, we also gained team collaboration skills as we initially struggled with communication and resolved these challenges

by providing feedback to each other and practicing open communication. We learned how best to communicate with one another, and we learned to be adaptable and incorporate feedback for a successful outcome which is this project that we hold here for you today.

**Potential Future Work:**

In our current database, demographic information such as race, ethnicity, gender, and socioeconomic status for both victims and perpetrators is not included. We recognize the importance of this missing data in providing a comprehensive understanding of the crime patterns. We identified using a free-use dataset that incorporates such demographic information to enhance, however adding that component is out of the assignment's scope. However, this would be a great step to increase the value of the system we created. By integrating this demographic data, our database would be a valuable resource for analyzing crime patterns across diverse communities, identifying disparities, and addressing socioeconomic factors that contribute to such criminal activities. This requires us to conduct thorough research to improve the database by ensuring the inclusion of crucial demographic information for a more insightful analysis of crime data.