**Project Title: Crime data Analysis**

**Introduction:**

Introduction: Our project aims to enhance community safety in Chicago through data analysis and technology.

**Significance:** Chicago faces challenges related to crime, which have been exacerbated by youth issues, and economic hardships. Community/Population Focus: Our primary focus is on the city of Chicago, with an emphasis on neighborhoods with high crime rates and social challenges.

**Clearly identify the community/population that is the focus of your project.**

**1. Community/Population Focus:**

Community/Population: The city of Chicago is a major urban center dealing with challenges. We believe our project can make a significant impact in addressing these challenges to our client Chicago police department.

**Identify the dataset(s) you will use including their source.**

**2. Dataset Description:**

Dataset(s): We will utilize multiple datasets, including crime data from the Chicago Police Department and demographic information from the U.S. Census Bureau. These datasets are publicly available.

**Sources:**

* Chicago Crime Data (provided): [Link to Dataset](https://www.kaggle.com/code/khalilabu/chicago-analysis-of-crime-data/input)﻿
* U.S. Census Data for Chicago: [Link to Dataset](https://data.cityofchicago.org/Health-Human-Services/Chicago-Population-Counts/85cm-7uqa)﻿
* Climate Data: [Link to Dataset](https://www.visualcrossing.com/weather-api?ga_api10&gclid=CjwKCAjwp8OpBhAFEiwAG7NaEudnP7uvDYhVpzuPkwyJDutghGpYXNQQS_Wclh9d7_65I6828cSJ4BoCstsQAvD_BwE)
* Sex offenders Data: [Link to Dataset](https://data.cityofchicago.org/Public-Safety/Sex-Offenders/vc9r-bqvy/data?no_mobile=true)
* Arrests Data: [Link to Dataset](https://data.cityofchicago.org/Public-Safety/Arrests/dpt3-jri9/data)

Source: The crime data is sourced from the Chicago Police Department, while demographic data is obtained from the U.S. Census Bureau.

**Illustrate what problems the dataset(s) will allow the community to solve.**

**3. Problem Solving:**

Problems: The datasets will enable the community, specifically the Chicago Police Department (CPD), to address limited resources, data accuracy, community trust, and legal/ethical concerns.

Solutions: By analyzing the data, we aim to improve resource allocation, enhance crime prevention strategies, and foster community trust. We will also adhere to ethical guidelines in data analysis and maintain data security.

**Identify any ethics or privacy concerns and how you will address them in your project.**

**4. Ethics and Privacy Concerns:**

Concerns: Potential privacy concerns may arise from handling crime data. Legal and ethical considerations in data usage are crucial.

Solutions: We aggregated data at a neighborhood level to ensure individual privacy. Workshops to client on data privacy and evidence handling will be conducted to ensure ethical data practices. We will plan to put views with access permission setting to desired people who are correct data owners.

1. Data Privacy and Protection - Implement robust encryption and access controls to safeguard sensitive information. Anonymize and aggregate data wherever possible to protect individual identities.

2. Transparency - Maintain transparency in how data is gathered, analyzed, and used in the project.

3. Non-Discrimination and Fairness - The project's outcomes and decisions are free from any sought of biases based on race, gender, age, or any other demographic factors.

4. Accountability and Responsibility - Regularly review and audit the system to identify and rectify any potential ethical concerns.

5. Views Instead of Tables:

* Opted for creating views instead of tables for sensitive information.
* Enhances security by restricting direct access to underlying data, providing an additional layer of protection.

6. Granular Access Permissions:

* Utilized database features to grant and revoke access selectively.
* Assigned permissions based on the principle of least privilege, ensuring individuals have access only to the necessary data.

**Why you chose the community/population.**

**1. Chosen community/population**

**Diverse Challenges:** Chicago faces a wide range of social and safety challenges, from high crime rates to economic disparities. This diversity of issues provides an opportunity to tackle a broad spectrum of problems through data analysis.

**Significance:** Chicago is one of the largest and most significant cities in the United States, making it an ideal location to address urban challenges. The impact of the project can be substantial due to the city's size and influence. There has been substantial economic loss due to the movement of companies to other cities and the tourism industry is significantly negatively impacted. All these reasons struck a spark for our interest in choosing this dataset which outlines our vision, methodology, and expected outcomes in to make Chicago a safer place for all its residents by helping the Chicago Police Department.

**How the dataset(s) you have chosen are related to the community/population you have chosen.**

**2. Dataset and Community Connection:**

Connection: The crime data relates to Chicago's crime challenges, allowing us to identify high-crime areas and trends. Demographic data provides insights into affected communities.

**Community Engagement:** A strong dataset-community connection can foster community engagement and support. When the community sees that the project is directly addressing their issues, they are more likely to get involved and support CPD’s initiatives.

**Trust:** Establishing this connection also builds trust with the community. It shows that our project is genuinely concerned with their well-being and is based on data that reflects their realities.

**Data-Driven Decision-Making:** By emphasizing the relevance of the dataset to the community, you encourage data-driven decision-making. This can lead to more effective policies, resource allocation, and crime prevention strategies.

**Advocacy and Awareness:** If the community understands how the data relates to their challenges, they are more likely to advocate for the results and use the findings to raise awareness about safety issues. This aligns with the project's goal of empowering media and activists.

**How you chose the questions and how they are related to the community.**

**3. Question Selection and relation to community:**

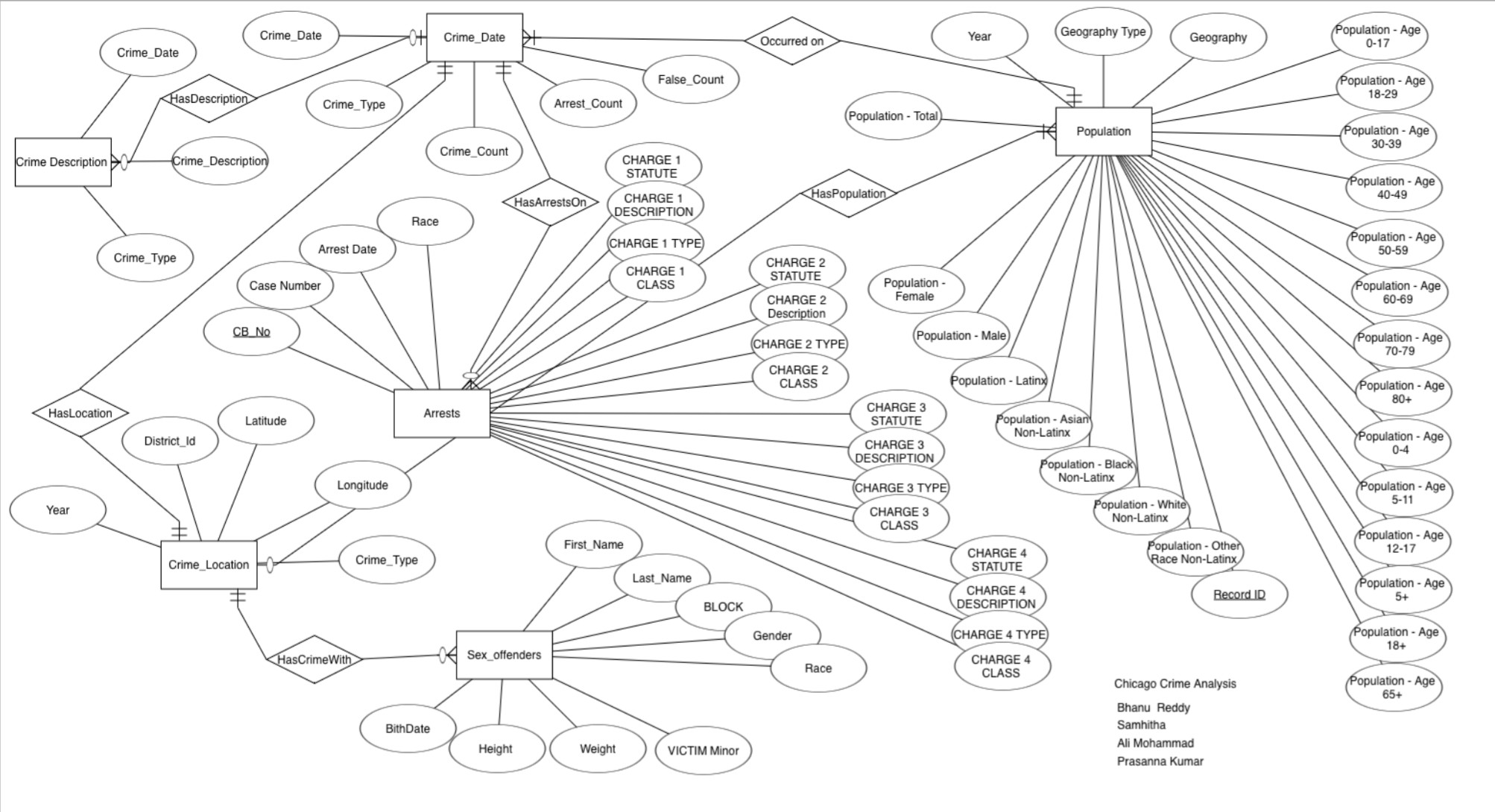
**Selection**: The questions have been chosen by consulting with the CPD as a client, understanding their priorities, and analyzing the data.

The questions we have chosen for our analysis relate to the Chicago community's safety, resource allocation, and decision-making processes. They are designed to address the pressing concerns of the city's residents and law enforcement agencies.

* How can we assess sex offender patterns to unveil the demographics of both male and female offenders, while simultaneously determining the number of minor victims involved in these crimes?
* Identify patterns in crimes occurring under different climatic conditions.
* Identify areas with a high concentration of crimes, enabling law enforcement to allocate resources and increase surveillance in these hotspots.
* Compare crime rates with population density in different areas to identify patterns
* Analyze trends in diverse types of crimes across specific time periods.
* How can we leverage offenders' data to gain insights such as age demographics to criminal activities
* Identify and analyze the top three primary types of crimes for each week of the year, considering various districts?

**Chen ER diagrams of your data illustrating how you will migrate your dataset(s) to a relational database.**

**4. ER Diagram and how we plan to migrate the data**



We intend to use the MySQL Workbench Data Import Wizard to migrate our datasets, which are currently in CSV format with tab-separated values, to a relational database. Once imported, we will refine the schema and data types to ensure alignment with our project's requirements and objectives. This process will enable us to structure the data optimally for analysis and to make informed decisions based on the dataset.

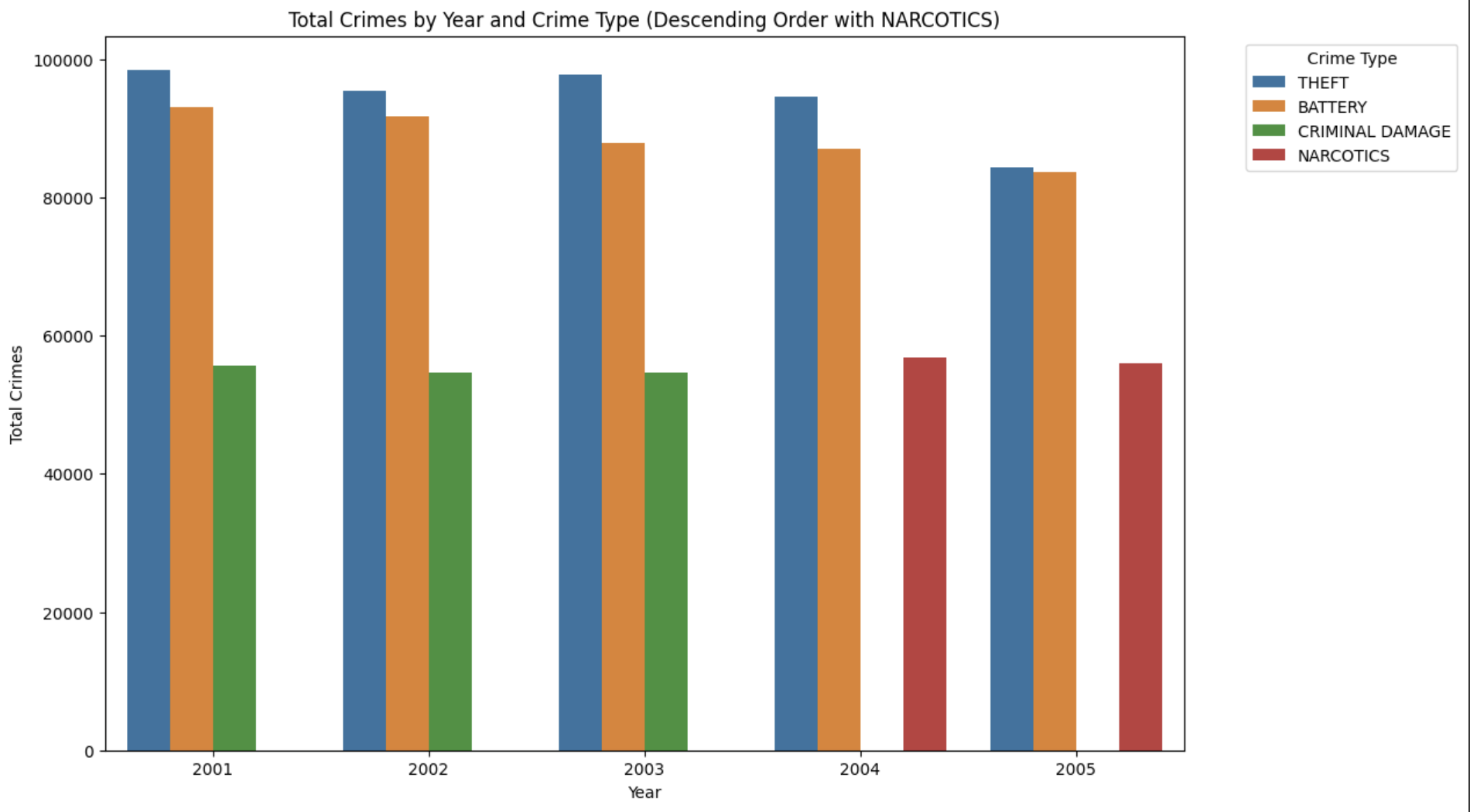
**Detailed report on analysis:**

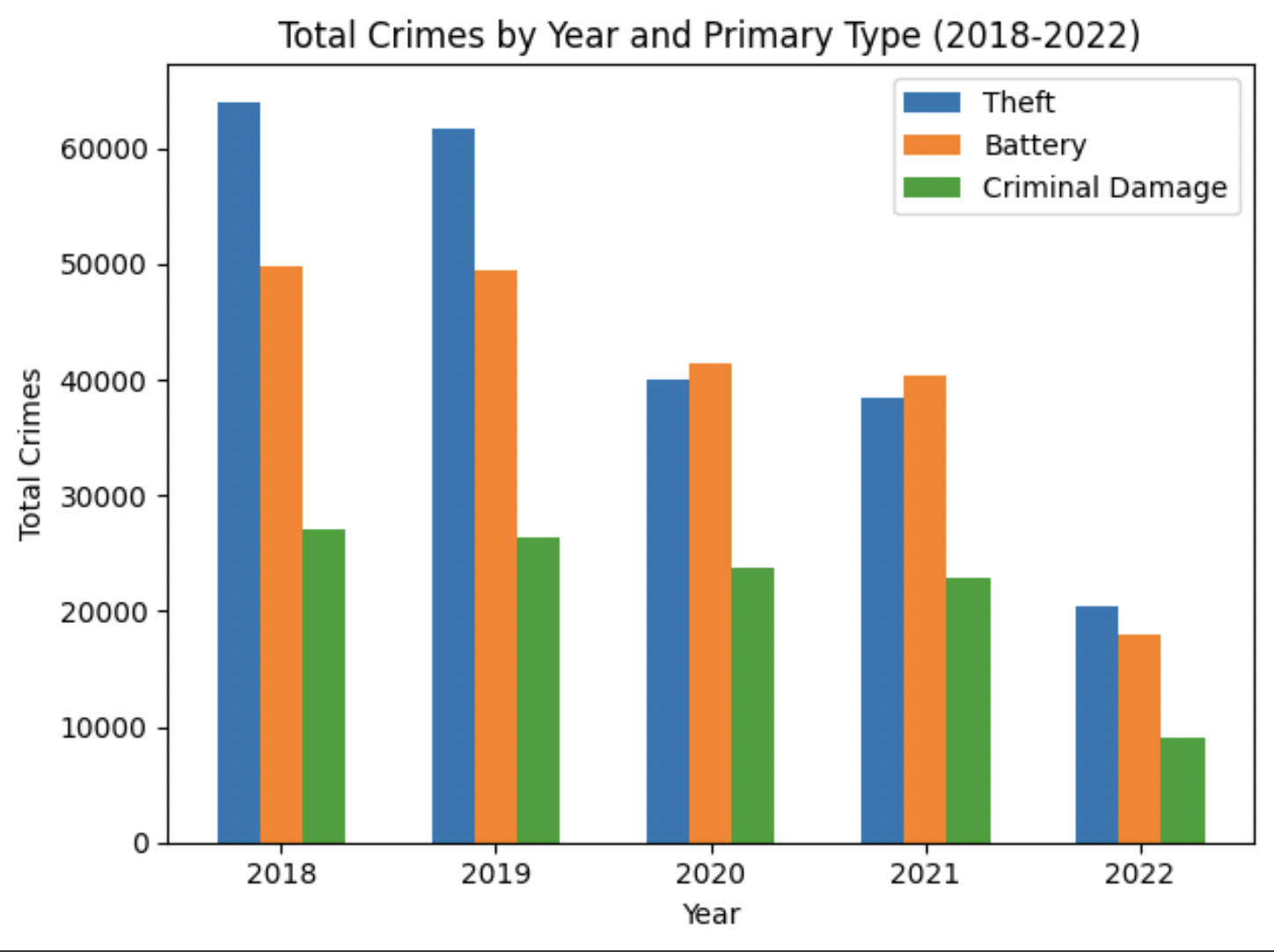
**Usecase: Analyze trends in different types of crimes across specific time periods .**

Our team addressed the need for comprehensive insights into crime trends from 2018 to 2022, focusing on identifying the top three crime types each year. The 'RankedCrimesView' was designed to streamline this analysis by utilizing SQL window functions and common table expressions (CTEs). This view facilitates a quick understanding of the most prominent crimes annually, offering valuable information for strategic decision-making and resource allocation by law enforcement agencies. The tiered ranking system enables a nuanced exploration of crime patterns, assisting in the identification of significant trends year over year.



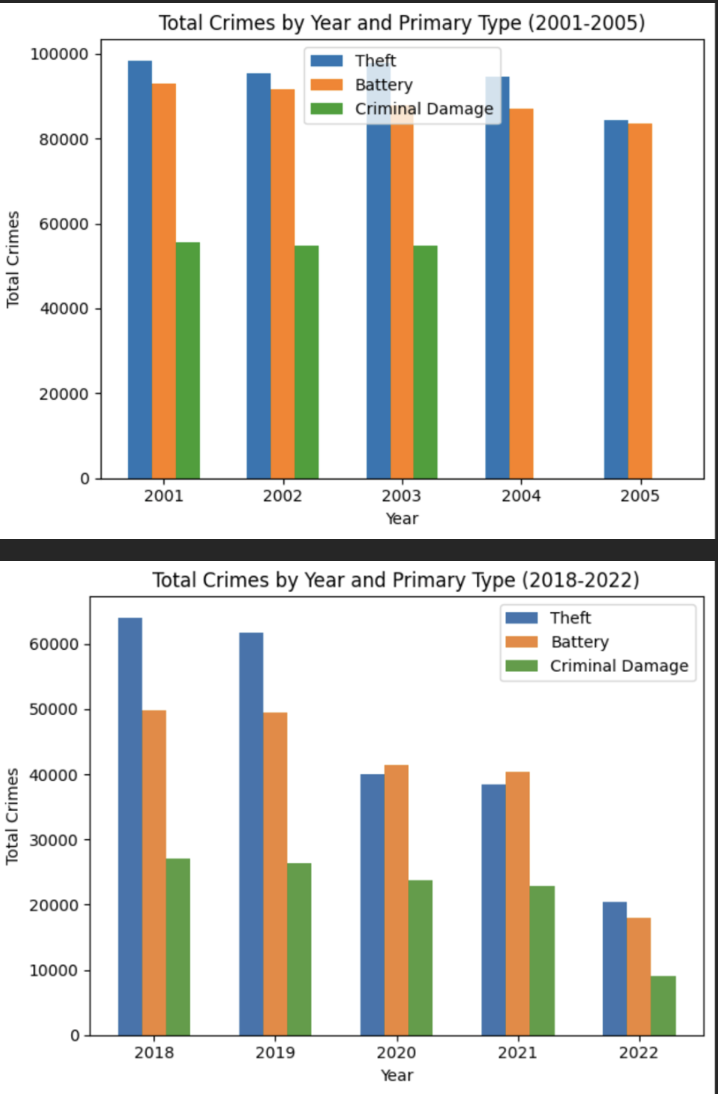
**Here we are trying to analyze the crime patterns over the years. We are analyzing the crimes between 2001-2005 and 2018-2022. This gives us an idea about the crime rate and different types of crime committed.**





The visual representation highlights a discernible trend showing a reduction in the overall number of crimes over time. In the early 2000s, theft, battery, and criminal damage ranked as the top three crimes. However, a notable shift occurred in 2004 and 2005, with narcotics emerging as one of the primary offenses in Chicago during those years. Examining more recent crime data reveals a continued decrease in the total number of crimes, suggesting that contemporary technology and effective policing efforts have contributed to heightened awareness and successful crime reduction strategies.

Python file link: <https://colab.research.google.com/drive/1rO7KZgJXq61jihV5BY6Dvj_c4uRQU_IN?usp=share_link>



Here we can notice the top three crimes as per data were “Theft, Battery and criminal damange” for years ranging from 2018 till 2022. We can also downward trend in count by year.

Furthermore, recognizing the importance of understanding year-over-year changes in crime rates, we developed the 'RankedCrimesYoYView.' This view not only provides the top crimes each year but also calculates the percentage change in crime rates from one year to the next. The 'Year\_over\_Year\_Change' CTE assesses trends, classifying them as 'Increase,' 'Decrease,' or

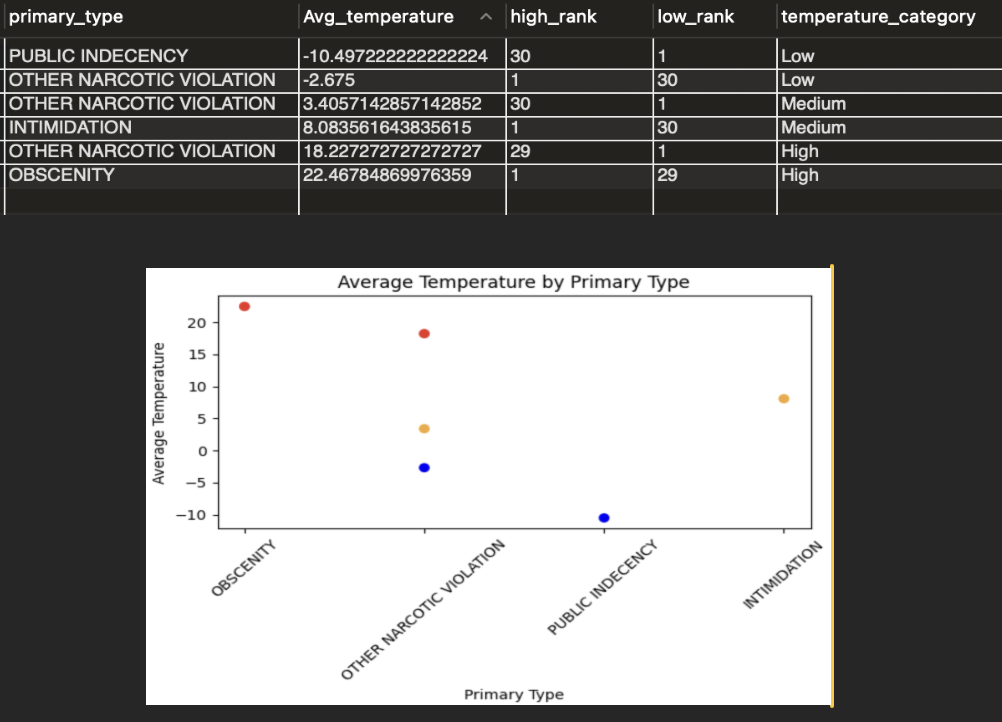
  **Yoy analysis of Crime trends**

'No Change.' This approach ensures a dynamic analysis of crime patterns, allowing for proactive adjustments in law enforcement strategies and community-focused initiatives. These views collectively empower stakeholders with actionable insights to address crime-related challenges and enhance public safety effectively.

**Usecase: Identify patterns in crimes occurring at different climatic conditions.**

In response to the use case aimed at uncovering potential correlations between crime rates and temperature variations, our SQL queries provide a detailed analysis. The initial query isolates the highest and lowest temperatures recorded in the climate data, serving as a preliminary overview of the temperature range under consideration. The subsequent query, encapsulated in the 'CrimeTypeByTempView,' delves deeper into this correlation. It employs a multi-step process, utilizing common table expressions (CTEs) to link crime data with temperature information. The script categorizes temperatures into three ranges—Low, Medium, and High—via a CASE statement.

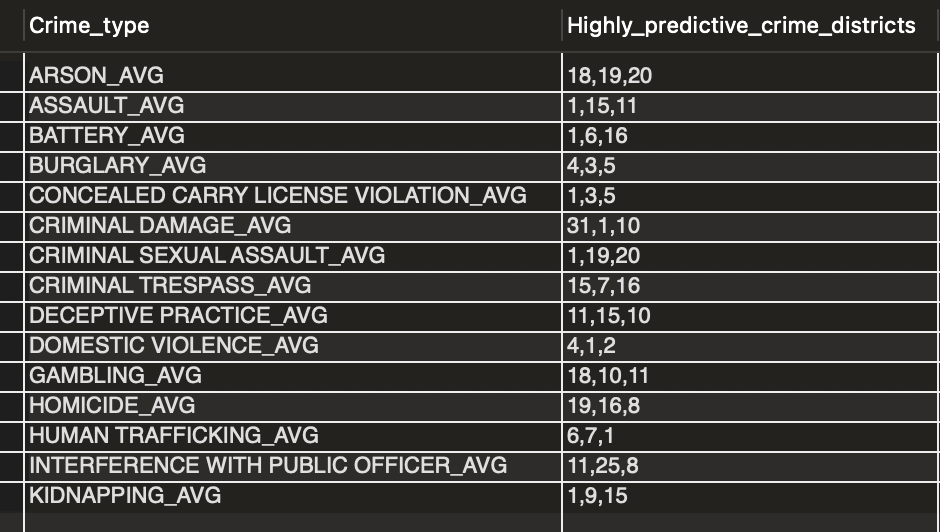
Basically for every category we see which aggreagated average temperature related to crime type falls close to border line temperatures. Public indecency will be have lease average temperatures and other narcotic violation has highest temperature with in the Low category when sorted and acts as highest and lowest temps. Similary we assess other crime types for other two categories as well.



The SQL code utilizes row number logic to categorize average temperatures into 'Low,' 'Medium,' and 'High' temperature ranges based on predefined conditions. For each category, the code ranks average temperatures of all crime types in ascending or descending order, ensuring that the temperatures closest to the borderline of each category are assigned the lowest and highest row numbers. The selection of rows with row number equal to 1 retrieves the entries with the lowest or highest average temperatures within their respective categories. This approach offers a nuanced and precise classification of average temperatures, facilitating a detailed analysis of temperature patterns associated with different crime types.

**Usecase:** **Identify areas with a high concentration of crimes, enabling law enforcement to allocate resources and increase surveillance in these hotspots.**

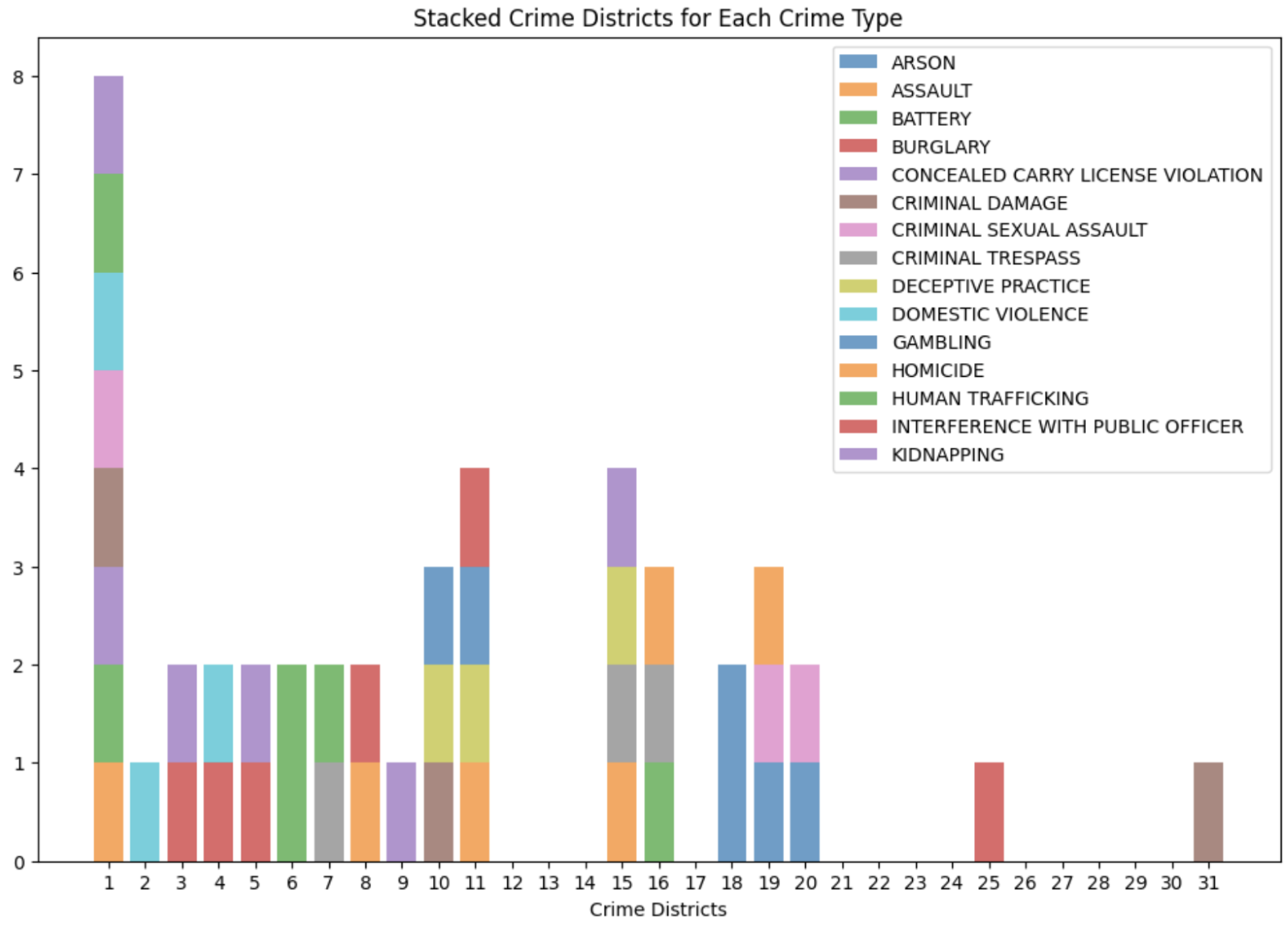
The SQL queries provided are designed to analyze and present crime data with a focus on districts, crime types, and year-over-year trends. The “CrimeTypeByDistrictView” offers a detailed breakdown of crime counts for various crime types within each district and year. By pivoting the data, the view provides a structured format that facilitates the analysis and comparison of crime trends across different districts over specified years. This structured presentation aids in identifying patterns and variations in crime rates.



Furthermore, the CrimeTypeByDistrictYoYView view delves into year-over-year changes in crime statistics. The calculated metrics, such as total crime count, lag values, percentage changes, and change status, offer a comprehensive overview of how specific crime types evolve over time for each district. This information is crucial for understanding and addressing fluctuations in crime rates, enabling informed decision-making for crime prevention strategies and resource allocation. Both views collectively contribute valuable insights into the dynamics of crime patterns, supporting evidence-based approaches to enhance public safety.

Additionally, the PredictedDistrictsByCrimeTypeView view identifies districts that are most likely to experience an increase in crime rates for specific crime types. By highlighting the top three districts with the highest average percentage change for each crime type, this view provides actionable information for targeted crime prevention efforts.

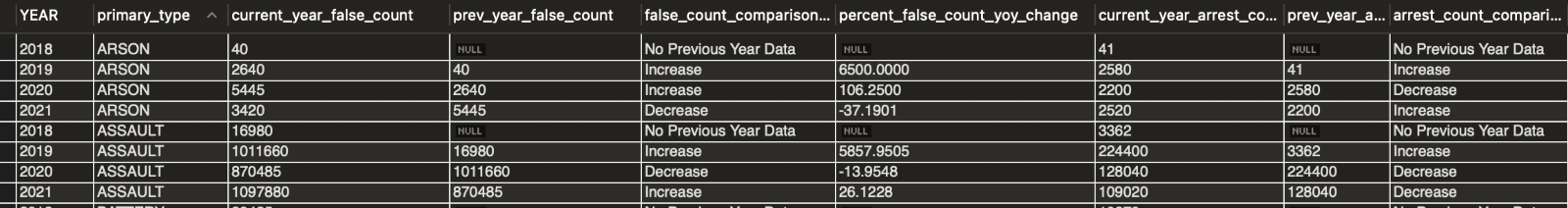
The concatenation of district names further streamlines the interpretation of districts prone to certain crime types. These views collectively empower law enforcement and policymakers to make data-driven decisions, ultimately contributing to more effective crime management strategies while deploying units for resource allocation by district.



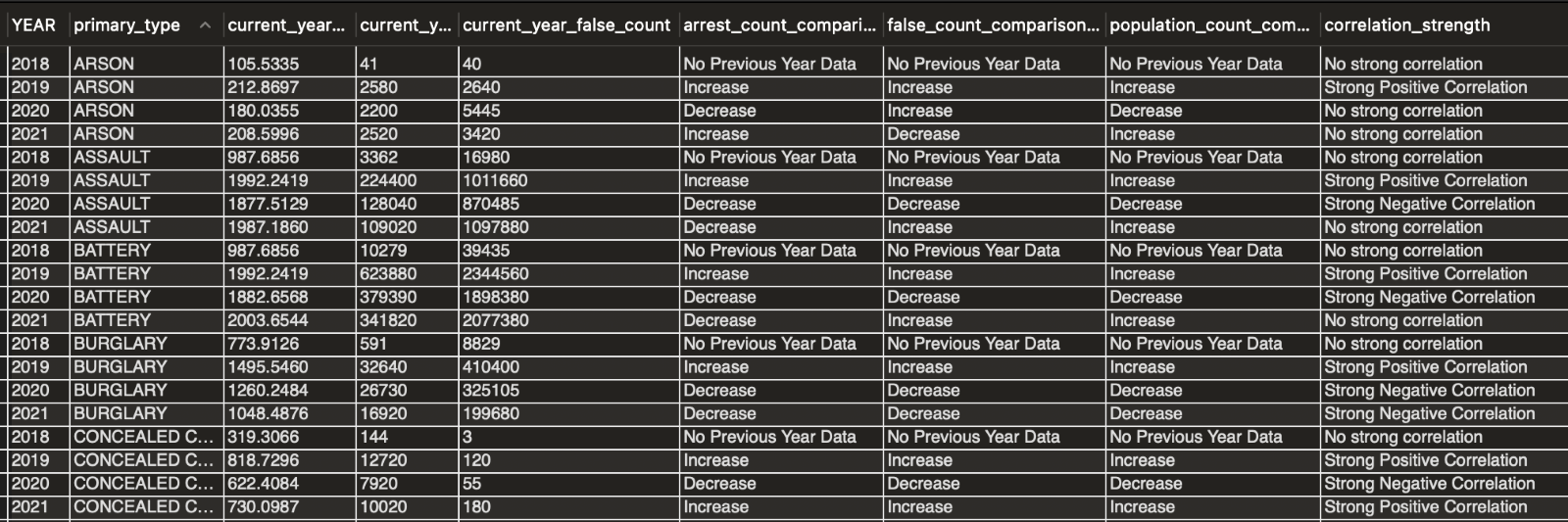
Here we can see nice visualization of how each district visualizes crime types that happened.

**Usecase: Compare crime rates with population density in different areas to identify patterns**

The SQL script facilitates the comparison of crime rates with population density by analyzing year-over-year changes in crime-related statistics (false count, arrest count) and population for various crime types.



The Common Table Expression (CTE) 'CrimeDataWithPreviousYear' efficiently aggregates the relevant data, incorporating LAG() functions to retrieve values from the previous year. The output categorizes changes as 'Increase,' 'Decrease,' or 'No Change,' providing valuable insights into the dynamic relationship between crime and population density. The subsequent analysis on correlation strength adds depth, allowing for a comprehensive understanding of how crime-related statistics correlate with population changes across different crime types.

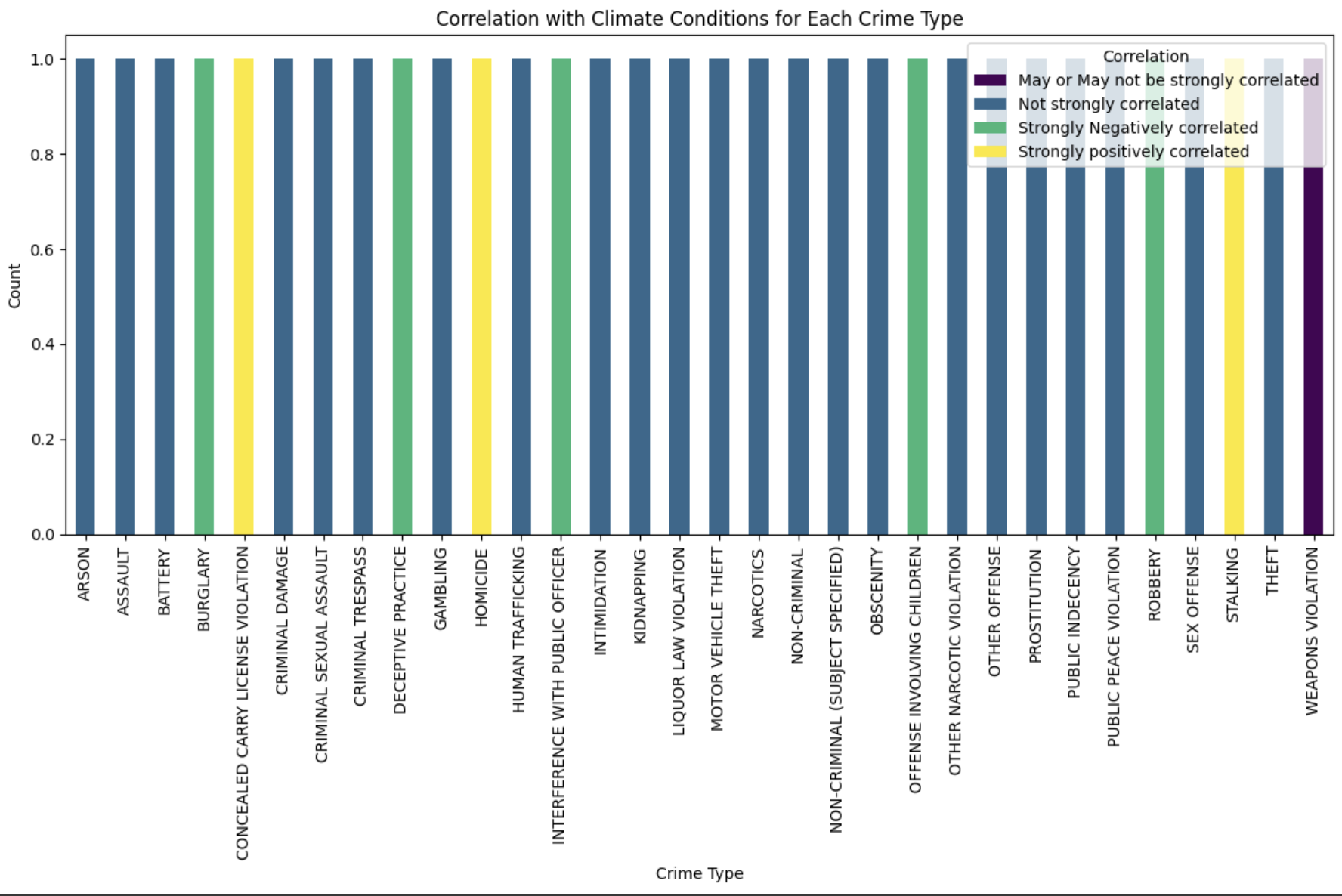




In the context of the project, this SQL code serves the purpose of identifying patterns in crime rates concerning population density.

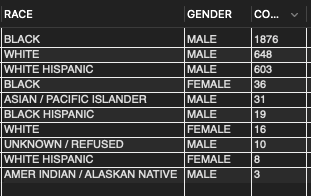
It offers a systematic approach to examining year-over-year variations, aiding in the recognition of correlations between crime-related metrics and population dynamics. The insights derived from this analysis contribute to a more nuanced understanding of how crime patterns evolve in relation to population density in diverse areas.

We compared arrest rate and population trend and basically assessed if population increase/ decrease had any effect on arrest rate and identified strong positive/negative correlation crimetypes with population trend.



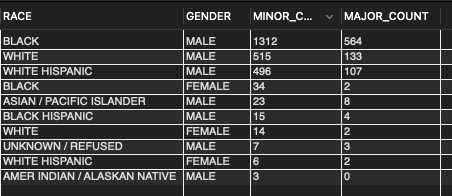
Use Case #: **How can we assess sex offender patterns to unveil the demographics of both male and female offenders, while simultaneously determining the number of minor victims involved in these crimes**

The observed patterns in the data distinctly reveal a predominant presence of male individuals among sex offenders. Additionally, it is noteworthy that the incidence of sex crimes committed by individuals of the black race is approximately three times higher compared to other racial groups.



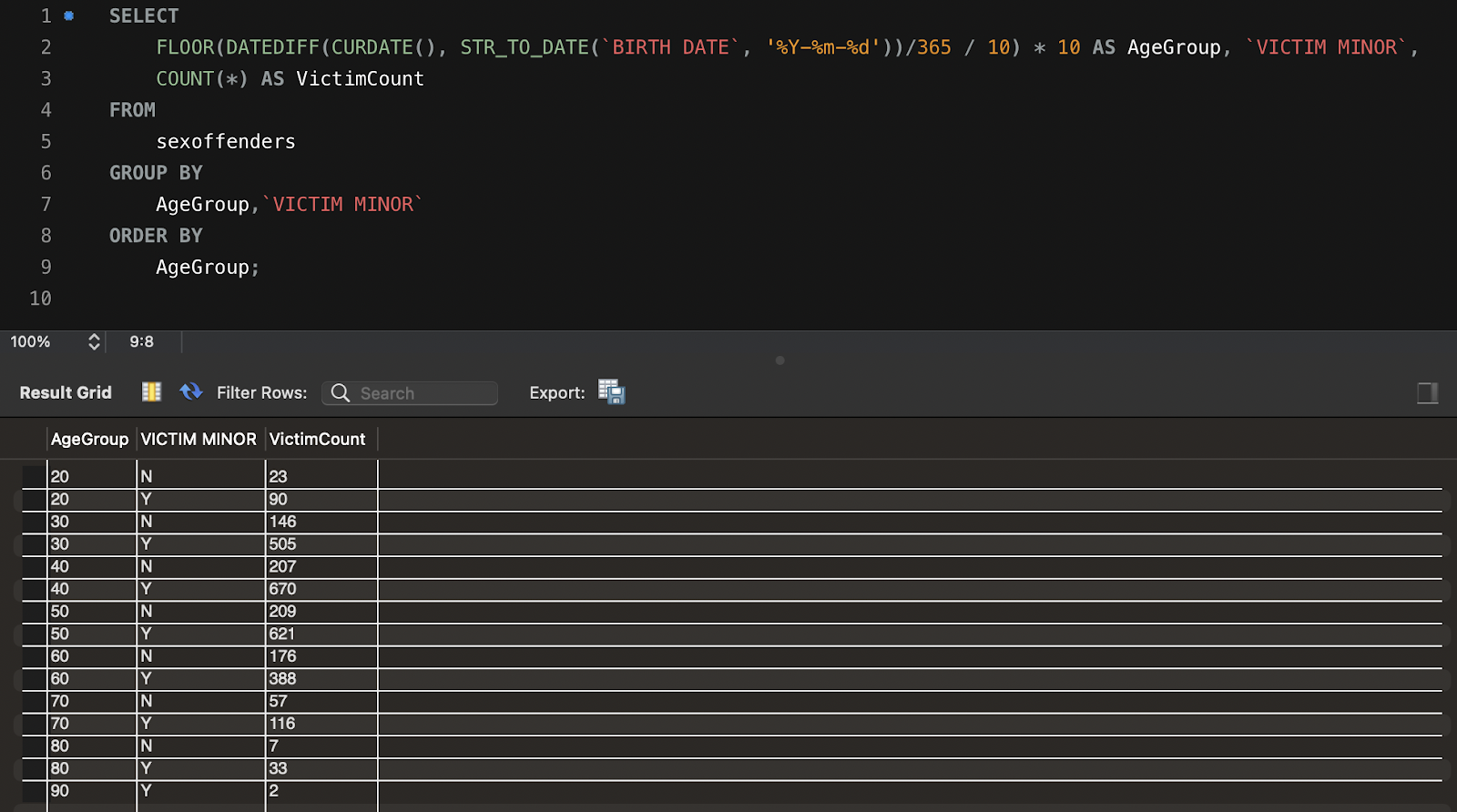
#To Find out number of Minor VICTIMS involved

Upon reviewing the outcomes, a notable trend emerges, indicating that the majority of victims were minors. The prevalence of minor victims is nearly twice as high when compared to the count of major victims.



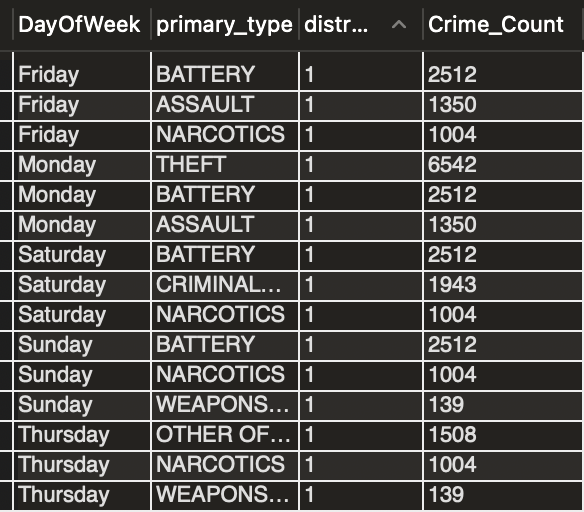
**Usecase: How can we leverage offenders data to gain insights such as age demographics to criminal activities:**

This Usecase has provided us a valuable insight for understanding the age dynamics associated with criminal activities involving sex offenders. By categorizing offenders into age groups and distinguishing between victims who are minors and those who are not, the query offers insights into patterns of criminal behavior across different demographic segments. The calculated age groups, structured victim data, and count information enable law enforcement and analysts to discern trends, potentially identifying areas of heightened risk and tailoring prevention strategies. Moreover, the ordered presentation of results allows for a clear visualization of how criminal incidents are distributed among age cohorts, contributing to a nuanced understanding of the intersection between offender age demographics and the nature of their offenses. Overall, this query facilitates a targeted and informed approach to addressing and preventing sex-related crimes by leveraging data-driven insights into the age characteristics of both offenders and victims.



**Usecase: Identify and analyze the top three primary types of crimes for each week of the year, considering various districts?**

Implementing these use case provided us a robust analytical framework for extracting meaningful insights from crime data within a temporal context. The first query, focusing on the day of the week, facilitates a detailed examination of crime patterns based on both the primary crime type and the district.

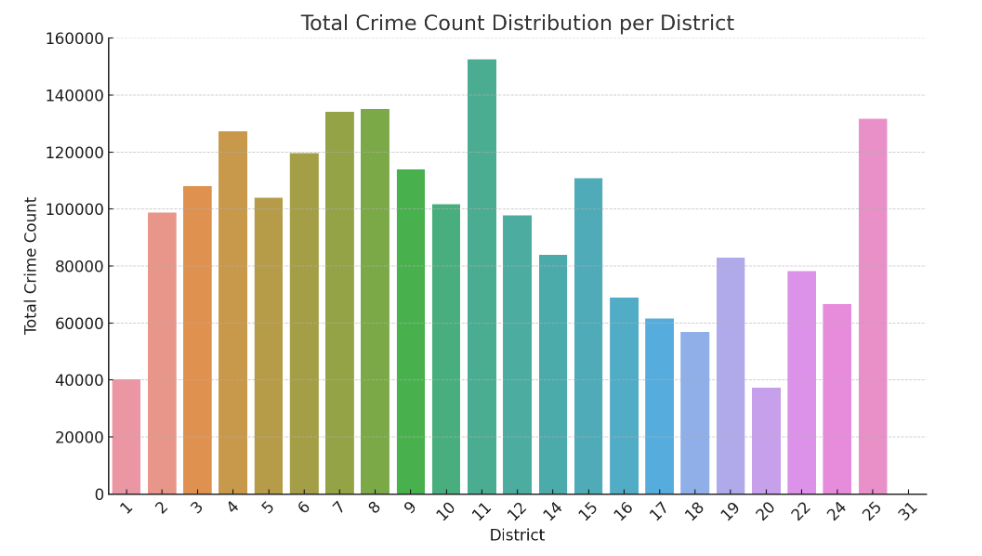


By identifying the top three crimes for each day and district, law enforcement and city planners gain a nuanced understanding of when and where certain offenses are most prevalent. This information can inform resource allocation, optimize law enforcement strategies, and guide community engagement initiatives to address specific crime trends effectively.

It appears that the query specifically highlights the prominence of 'NARCOTICS'-related crimes across various districts on Fridays, which could indicate a trend that may require targeted interventions.



The second query, centered on the week of the year, extends the temporal analysis to a broader scale. Investigating the top three crimes for each week with district information provides a more macroscopic view of crime trends over time. This can be invaluable for detecting seasonal variations or long-term patterns in criminal activities.





The ordered presentation of results allows for a clear visualization of how crime incidents evolve throughout the weeks, enhancing the decision-making process for law enforcement and city planners. In conclusion, these SQL queries empower stakeholders to extract actionable intelligence from crime data, fostering data-driven strategies for crime prevention, resource optimization, and community safety.

**Database Management System Optimization Report**

Our project aimed to address key challenges in the MySQL Workbench environment, focusing on the Partition window limitation and handling schema changes post-load. Additionally, we implemented best practices to enhance security, maintainability, and comprehension of the database structure.

**Challenges and Solutions:**

* MySQL Workbench Partition Window Limitation:
  + Identified a limitation in the Partition window with regards to the number of columns. We have limit the output columns to counter issue.
* Handling Schema Changes Post Load:
  + Recognized the need for handling schema changes efficiently, especially with timestamp columns.
  + Developed a solution by standardizing input formats and transforming them into a custom standard output format.

Best Practices Implemented:

* View Creation for Unauthorized Changes Prevention:
  + Strategically utilized views to prevent unauthorized changes to critical database structures.
  + Ensured data integrity and security through controlled access to underlying tables.
* Fine-Grained Access Control with GRANT and REVOKE:
  + Employed GRANT and REVOKE statements to grant specific access privileges to authorized users.
  + Established a robust access control mechanism to enhance security and restrict unnecessary actions.
* Utilization of Common Table Expressions (CTE):
  + Integrated Common Table Expressions (CTE) for improved query readability and comprehension.
* Lag Function Implementation for Year Over Year Comparison:
  + Implemented the Lag function to perform year-over-year comparisons efficiently.
  + Enhanced analytical capabilities for trend analysis and decision-making processes.

Conclusion: In conclusion, our project successfully addressed critical challenges within the MySQL Workbench environment and implemented best practices for database management. By optimizing the partitioning process, handling schema changes effectively, and incorporating security measures, our approach has significantly improved the overall functionality and maintainability of the database system.

**Conclusion:**

Conclusion: We are excited to present our findings and recommendations, which we believe will make a significant impact in addressing the identified pain points. Our project aligns with the CPD's objectives, and the information provided in this proposal. We look forward to contributing to safer communities.