# 

# **BY : RAVNEET SINGH**

Problem Statement

A year ago in Chicago, the high incidence of crime prompted the government to take decisive action by launching an operation with top-performing police officers. As a year has passed since this decision, the government seeks to evaluate the reduction in crime cases and the advantages gained from this new operation. The government has engaged you as a freelancer data analyst to analyze the changes and improvements in crime rates.  The government is seeking the following objectives from your analysis:

1. If the case-solving time improved
2. What are the localities where the crime rate was higher
3. What can be the measures that can be taken to improve the crimes further

Chicago has grappled with high crime rates, which have posed significant challenges to the safety and well-being of its residents. Recognizing the urgency of the situation, the government took decisive action to mitigate crime and restore confidence in the city's security infrastructure. The implementation of this operation marked a pivotal moment in Chicago's efforts to combat crime and improve overall quality of life for its residents.

**Primary Objectives:**

1. **Assess Improvement in Case Solving Time:** Determine whether the new operation has led to a reduction in the time taken to solve criminal cases, reflecting enhanced efficiency and effectiveness in law enforcement procedures.
2. **Identify High-Crime Localities:** Analyze crime data to identify areas within Chicago where crime rates have historically been higher, highlighting priority areas for targeted interventions and resource allocation.
3. **Recommend Crime Control Measures:** Propose actionable measures and strategies to maintain control over crime rates moving forward, leveraging insights gained from data analysis and best practices in law enforcement.

**Objective Questions**:

1. In analyzing the provided dataset with Power BI, ensure data cleaning to address inconsistencies and missing values before further analysis.

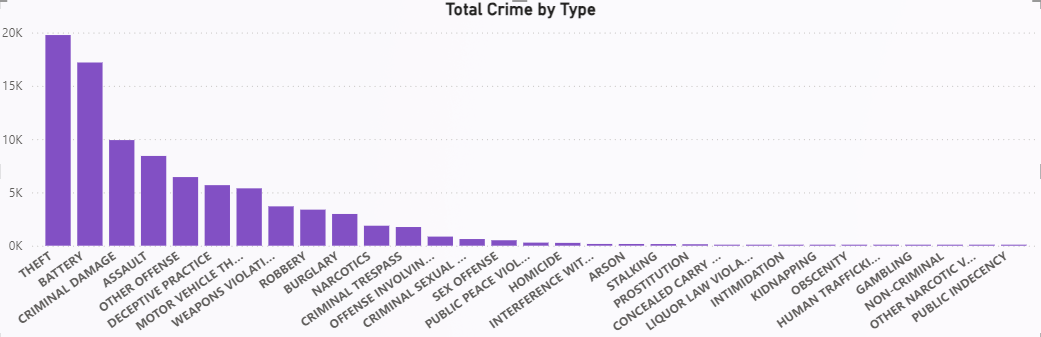
**Handle Missing Values:**

Depending on the context, decide how to handle missing values. Options include:

* + Resolve Case Time: Calculated the date difference between the Date and Updated columns using the formula Resolve Case Time = DATEDIFF('crimes data 2022'[Date], 'crimes data 2022'[Updated On], DAY).
  + Month: Extracted the month from the Date column using the formula Month = FORMAT('crimes data 2022'[Date], "MMMM").
  + Latitude and Longitude: Created using the Location column through the "Add Column by Example" feature.
  + Some columns in this data like Ward, X-coordinate & Y-coordinate, date format

1. Crime Type Analysis: Assess the frequency of each crime type to identify the most prevalent crimes occurring in the area.

Upon completion of the analysis, we will identify the top crime types based on their frequency of occurrence. This information will shed light on the predominant criminal activities in Chicago, which may include offenses such as theft.



1. Arrest Rate Evaluation: Analyze the percentage of reported incidents that have resulted in an arrest to gauge law enforcement effectiveness.

The calculated arrest rate provided valuable insights into the performance of law enforcement agencies in addressing reported incidents.

First, I have calculated the total crime measure, total arrest measure and arrest rate %

Arrest Rate % = 12.44

Used DAX Function for calculating arrest rate %

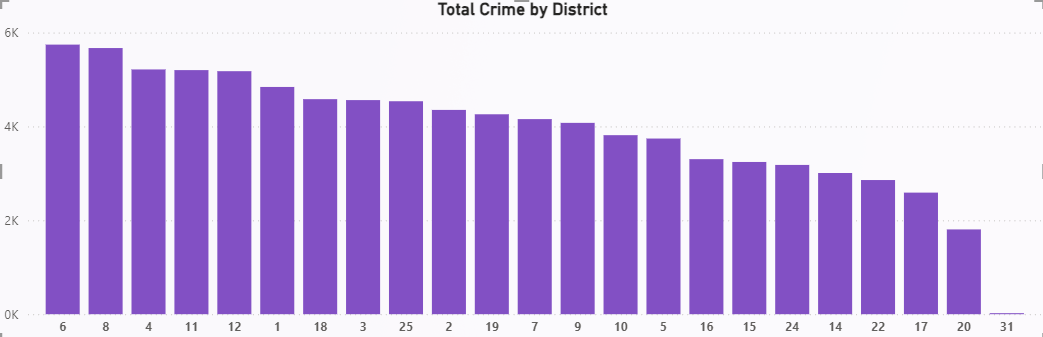
Arrest Rate (%) = DIVIDE([Total Arrests], [Total Crime]) \* 100



1. **District Crime Distribution Assessment: Calculate the number of crimes in each district to understand how crime is distributed across the city and identify high-crime areas.**

District crime distribution, we would perform with clustered column chart in which we select count to district and total crime and we found that

* District 6 is the 1st highest crime area with the number of 5733.
* District 8 is the 2nd highest crime area with the number of 5664



1. **How many categorical attributes are there in the data?**

In the dataset, there are 13 categorical attributes, representing qualitative information such as demographic characteristics & crime type. These attributes provide essential context for understanding patterns and trends within the data.

1. **Were there any Null values in the data, if there were how did you handle them? What is the ideal way to handle Null values?**

* I observed 944 null values in both the X-coordinate and Y-coordinate columns, constituting approximately 1% of the total dataset.
* To maintain data integrity and considering the small proportion, I decided to remove these null values, as their absence is not expected to significantly impact the overall dataset.

1. **Domestic Crime Proportion Analysis: Analyze the ratio of domestic-related crimes to other types of crimes to understand the prevalence of domestic incidents**.

Understand the prevalence of domestic Incidents we have to create a new measure with the name of domestic crime ratio

First, we have calculated the domestic crime measure and divide it with the total crime.

Domestic Crime = CALCULATE(COUNTROWS(Crimes\_data\_2022),Crimes\_data\_2022[Domestic]=True)

Used DAX Function for calculating domestic crime ratio

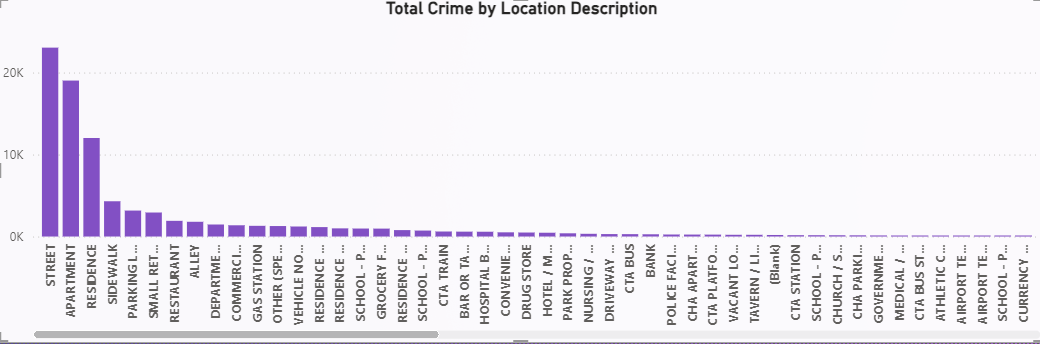
Domestic Crime Ratio = DIVIDE([Domestic Crime],[Total Crime])



1. **Is there any “Location Description” where the number of crimes is higher than expected? Come up with a table or visualization in which one can judge the frequency of crimes at each Location Description type**

Location crime distribution, we would perform with Clustered column chart in which we select count to location description column and total crime and we found that.

* In location description **street** has the 1st highest number of crimes which is 23039.
* In location description Column **apartment** has the 2nd highest number of crimes which is 19022.
* In location description Column **residence** has the 3rd highest number of crimes which is 12018.

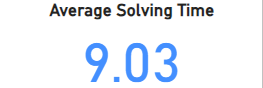
****

1. **What is the average time between reporting and solving a case as per the data?**

We can use the DATEDIFF function to calculate the difference between two dates. We will calculate the difference between incident occur date and closed date of incident.

* First, we have to create a new column (Average Solving Time) and use DATEDIFF function to calculate.
* After that in the report view, we choose card to show average no. of that column.

Please Check my power BI Main Tab for reference.



1. To reward the patrol officers, find the patrol area where the crimes reported were under control.

Analyse the crime data to identify patrol areas with low crime rates or where reported crimes are consistently decreasing over time. We find the beat column for patrol areas.

We have to create a visual table with total crime and beat:

* In beat no 1655 has the 1st lowest number of crimes which is 55.
* In beat no 1652 has the 2nd lowest number of crimes which is 57.
* In beat no 1654 has the 3rd lowest number of crimes which is 90.

1. Did you create any calculated columns in this project? What is the difference between the ‘calculated column’ and ‘add column’ functions?

Yes, I did create calculated columns in this project. Additionally, I utilized the

'Add Column' feature in Power BI Query to create three additional columns:

1. **Longitude and Latitude:** I generated these columns by using the "Column from Example" feature, where I provided sample data separated by a comma as the delimiter. This allowed Power BI to infer the pattern and create the Longitude and Latitude columns accordingly.
2. **Moment of Register Case:** Using the 'Extract Last 2 Characters' function, I extracted the last two characters from the date column to determine the moment of registering each case. This provided insights into the timing or phase of case registration within the month.

In total, I created 8 calculated columns using DAX formulas, which included operations such as arithmetic calculations, logical comparisons, and string manipulations. These calculated columns augmented the dataset with additional derived information, facilitating deeper analysis and insights.

1. Using ‘Calculate’ and a row iteration DAX function calculate the number of crimes which are of type ‘theft’ and happened in ‘District 8’.

Using DAX function we calculated crime type – Theft & District – 8 is 1143.

Used DAX function is:

Crime type "Theft" & District "8" = CALCULATE(COUNTROWS(Crimes\_data\_2022),FILTER(Crimes\_data\_2022,Crimes\_data\_2022[Type]="Theft" && Crimes\_data\_2022[District]=8))

**Please check 1Measure for reference**

1. Using Power BI can you separate the Longitude and Latitude from the Locations Column (Longitude, Latitude)? Which feature will you use?

Yes, in Power BI, we can separate the Longitude and Latitude from the Locations column using the 'Split Column' feature in Power Query Editor.

1. When we add a column in Power Query what’s the code that comes in M language in formula bar? What do you know about M-query?

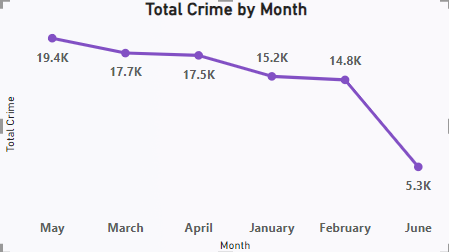
M-query, also known as the Power Query Formula Language, is a functional language used for data transformation in Power Query. It allows users to define step-by-step transformations on their data within Power Query Editor. M-query supports a wide range of operations such as data type conversion, splitting columns, merging tables, filtering rows, and more.

With M-query, users can perform complex data transformations and data cleaning tasks to prepare their data for analysis. It provides a powerful and flexible framework for manipulating data within Power Query, enabling users to create custom data transformation processes tailored to their specific requirements.

**Subjective Questions:**

1. **Is there any month-wise change in crime rates? If not, what could be the mistake in that operation?**

We are preparing visual line chart to show the monthly basis trend.

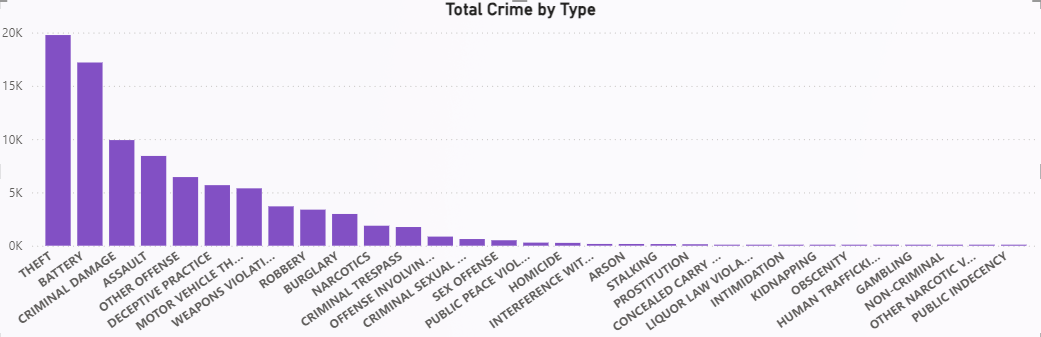


In analyzing the month-wise change in crime rates from January to June 2022, several key observations were made:

* A gradual increase in crime rates was noted until May, followed by a significant decrease in June. This drop could be due to the limited data available only up to the 9th of June.
* May had a recorded crime rate of 19,353 incidents, making it a month with a higher-than-average crime rate. This necessitates a deeper investigation into the factors behind this surge.
* February had a higher arrest rate of 13.83%, indicating a peak in law enforcement efficiency. Understanding the dynamics behind this increase is crucial.

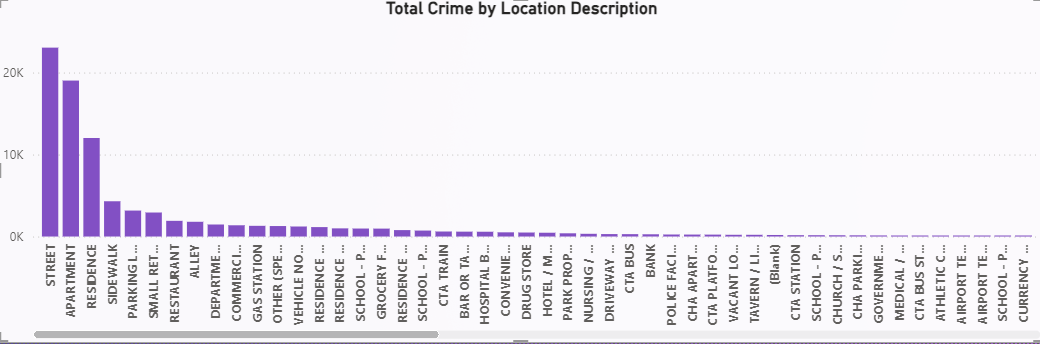
By focusing on the factors contributing to May's surge, strategies can be developed to reduce overall crime in Chicago.

1. How can we reduce the no. of crimes, and which types of crime should we focus on to achieve improvement in the overall number of crimes?



The increase in theft and battery crimes suggests a need for targeted interventions such as increased police patrols in hotspots, community engagement programs to prevent theft, and educational campaigns on conflict resolution to reduce battery incidents. Utilizing data-driven approaches and collaboration with law enforcement agencies can help develop effective strategies to address these specific crime types and improve overall public safety.

1. **Which localities experience higher crime rates, and what measures can we ensure to reduce these numbers?**

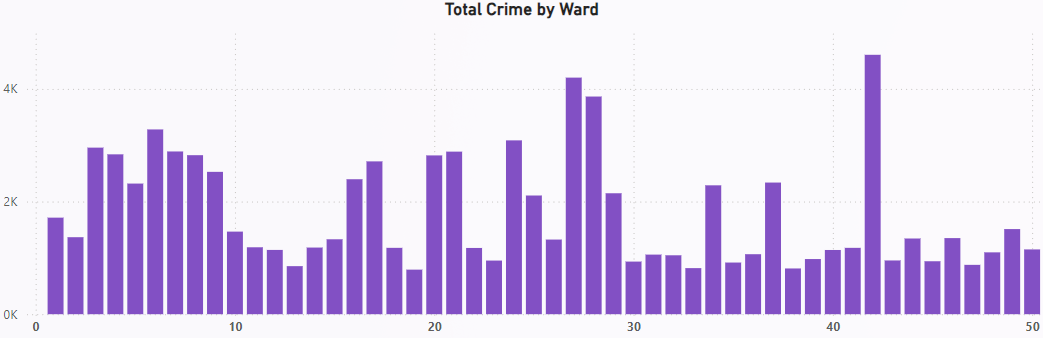
****  
﻿

* Total Crime was highest for STREET at 23039, followed by APARTMENT and RESIDENCE.﻿
* STREET accounted for 25.68% of Total Crime.﻿

**Insights**

To reduce the high crime rates in street, apartment, and residence locations, strategies like improving lighting and visibility, installing surveillance cameras, implementing neighbourhood watch programs, and collaborating with local authorities for regular patrols and rapid response can help deter criminal activities and create safer environments.

1. **Can you suggest wards where security improvements should be made to reduce crime?**



* Ward 42 has the highest crime rate with 4605 reported incidents.
* Ward 27 follows closely as the second highest with 4200 reported incidents.
* Ward 28 ranks third with 3864 reported incidents.
* Targeted strategies and interventions are necessary in these wards to address crime trends and enhance public safety.
* We have to improve security of these mentioned wards to reduce crime.

1. **Crime Rate Trend Analysis: Monitor changes in crime rates over time to detect any discernible patterns or trends.**

Solution:-

The analysis of crime data for the year 2022 reveals variations in crime rates across different months, with May showing a notable peak. Additionally, a weekly trend analysis highlights Tuesdays and Saturdays as peak days for criminal activities.

While this snapshot provides valuable insights, it's essential to acknowledge the limitations of analyzing data from a single year. To detect discernible patterns or trends accurately, it is advisable to collect and analyze data spanning multiple years. Longitudinal data would enable a more thorough examination of fluctuations, seasonal variations, and long-term trends in crime rates.

From a location perspective, we observe that crime is evenly distributed across various areas, making it challenging to discern any specific trends or patterns. However, a significant disparity emerges when analyzing crime occurrence during different times of the day. Crime rates notably increase during nighttime hours, indicating a need for more proactive measures during this period.

Addressing nighttime crime requires proactive efforts from authorities, such as increased patrolling, enhanced lighting in vulnerable areas, and community engagement initiatives tailored to nocturnal activities. By focusing resources and attention on nighttime crime prevention strategies, authorities can effectively reduce overall crime rates in the city and enhance public safety.

1. **Create a month wise tabular data consisting of two columns, month and total no. of crimes in that month. Also, add one more column where each row of the column contains the total no. of crimes for the previous month. Do we need to use any filter-based DAX function here (All, All except, etc)?**

Solution:

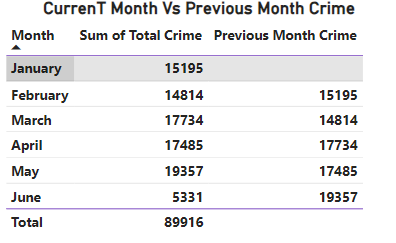
For my Power BI project, I encountered a scenario where I needed to analyze month-wise crime data. To accomplish this, I first created a continuous date table to ensure that I could utilize DAX functions effectively. However, due to the requirement for calculating the total number of crimes for the previous month, I realized the need for a different approach.

I subsequently created a new table named "Dates\_crime," which contains all distinct dates associated with the crimes that occurred on those specific dates. Then, I established a one-to-one relationship between this new table and my main date table. This relationship allowed me to accurately calculate the total number of crimes for the previous month using the PREVIOUSMONTH function in DAX.

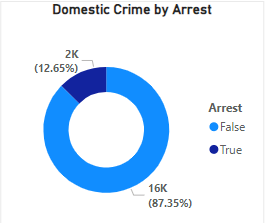
To visualize the relationship between the current and previous months' crime data, I deployed a table in Power BI. This table enabled me to compare the total number of crimes for each month with those of the preceding month. Through this analysis, I identified significant differences in crime rates between certain months.

Overall, by structuring my data model in this manner and utilizing DAX functions effectively, I was able to gain valuable insights into month-to-month variations in crime occurrences, which are essential for making informed decisions and implementing targeted interventions.

Dax Formula -> previous month = CALCULATE(SUM(Dates\_crime[Total Crime]),PREVIOUSMONTH('Dates Table'[Date]))

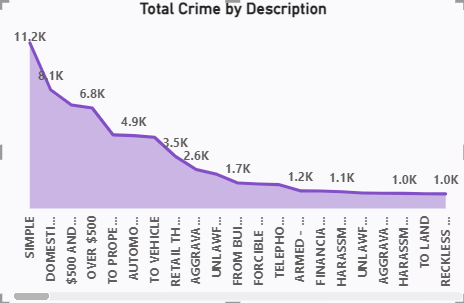


1. **As per the previous reports, most domestic crimes do not result in arrest due to public hesitation and family pressure, is this trend also visible in our data?**



* The data shows a higher number of domestic crimes reported for "False" (15725) compared to "True" (2277), indicating a significant disparity in incidence rates.
* "False" cases accounted for a substantial proportion, representing approximately 87.35% of all domestic crimes, highlighting the prevalence of non-domestic-related incidents.
* A key factor contributing to the low arrest rate for domestic crimes could be public hesitation and family pressure, which may deter victims or witnesses from reporting or cooperating with law enforcement.
* Addressing public awareness, providing support services for victims, and strengthening community outreach programs are essential to encourage reporting and improve law enforcement's effectiveness in addressing domestic crimes.
* Collaborative efforts between law enforcement, community organizations, and support services are crucial in creating a safer environment and reducing the prevalence of domestic-related incidents.

1. **Could you generate a visual representation that emphasizes the frequently occurring terms within the "Description" column?**

****

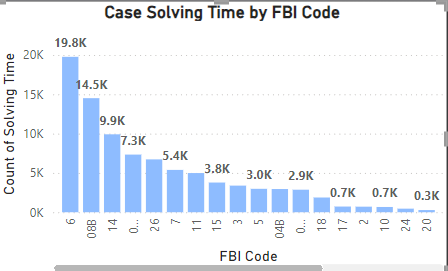
The analysis of the description column reveals several frequently occurring terms, including 'Simple' with 11,244 occurrences, 'domestic Battery Simple' with 8,060 occurrences, '$500 And Under' with 7,024 occurrences, and 'over $500' with 6,813 occurrences, among others. This visualization effectively highlights these common terms, offering insight into prevalent types of incidents.

By referencing this visualization, stakeholders can gain detailed insights into the distribution of frequently occurring terms. It's important to note that the visualization includes a filter, allowing users to focus on specific categories. Removing the filter expands the view, providing a comprehensive overview of term frequencies across all categories

1. **Are there any particular regions as per the data where the number of domestic crimes reported is very high?**

* The data does not include specific information about regions associated with domestic crimes, but it indicates a substantial overall number of domestic incidents reported.
* Analysing the geographical distribution of domestic crimes can offer insights into areas with higher prevalence rates and aid in targeting interventions effectively.
* Utilizing additional geographical data or mapping techniques can enhance the understanding of regional variations in domestic crime rates.
* Implementing targeted law enforcement strategies and community outreach programs in high-prevalence areas can help address domestic violence more effectively.
* Collaborating with local authorities, support services, and community organizations is crucial for developing comprehensive approaches to combat domestic crimes in specific regions.

1. **Is the solving time of cases also dependent upon the type and locality of crime?**

****

The type of crime & Locality can directly impact the complexity and duration of case-solving efforts

* Crime type complexity influences the duration and resources needed for case-solving efforts, affecting investigative strategies and timelines.
* Locality factors such as resource availability, investigative techniques, and community cooperation significantly impact case-solving outcomes.
* Analysing crime type and locality data provides valuable insights for optimizing law enforcement strategies and resource allocation.
* Enhanced community engagement and support are crucial in improving case-solving outcomes and fostering trust between law enforcement and communities.
* Tailoring investigative approaches based on crime type and locality considerations can lead to more efficient and effective law enforcement practices.

1. **Create a calculated column to flag the domestic crimes that took place in District 8?**

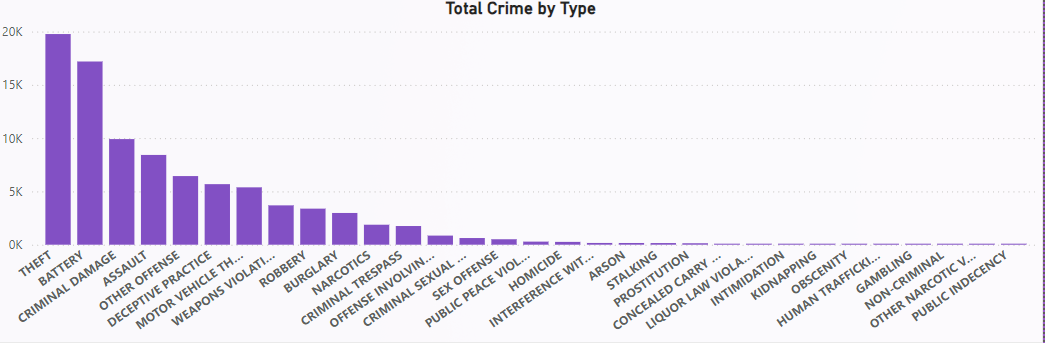
We have created calculated column "Domestic Crime in District 8" in Power BI:

Using DAX function:

DomesticCrimeDistrict8 = IF('Crimes\_data\_2022'[District] = 8 && 'Crimes\_data\_2022'[Domestic] = True, 1, 0)

* The calculated column flags domestic crimes that occurred in District 8 with a value of 1.
* Other crimes that do not meet these criteria are flagged with a value of 0 in the column.
* This column allows for easy analysis of domestic crime trends specifically in District 8.

1. **Out of all the types of crimes which do you think is the most dangerous one and rank the type of crimes according to their no. of occurrences?**

****

Ranking of crimes from most to least dangerous (theft, battery, criminal damage, and assault):

* Theft: While not physically violent, theft can have significant financial and emotional impacts on victims, especially if it involves high-value items or personal belongings.
* Battery: Battery involves physical violence or harm against another person, making it a more dangerous crime due to the potential for immediate injury or trauma.
* Criminal Damage: Criminal damage can result in property destruction or vandalism, posing risks to public safety and causing economic losses.
* Assault: Assault encompasses various degrees of physical violence and aggression, ranging from minor injuries to severe harm, making it one of the most dangerous crimes due to its potential for serious injury or even death.

1. **What do you understand by PowerBI gateway? What are its use cases?**

Power BI Gateway is a tool provided by Microsoft that serves as a bridge between on-premises data sources and the cloud-based Power BI service. It enables Power BI to access and refresh data from on-premises sources securely. Here's a breakdown of its key aspects and use cases:

* **Data Connectivity:** Power BI Gateway allows you to connect to various on-premises data sources such as SQL Server, Oracle, SharePoint, and more. It establishes a secure connection between these data sources and the Power BI service in the cloud.
* **Data Refresh**: One of the primary use cases of Power BI Gateway is to enable scheduled data refresh for on-premises data sources. It allows Power BI reports and dashboards to stay up-to-date by automatically refreshing data at regular intervals.
* **Direct Query:** Power BI Gateway supports Direct Query mode, where queries from Power BI reports are sent directly to the on-premises data source in real-time. This enables users to analyze the most current data without storing it in the cloud.
* **Personal and Enterprise Gateway:** Power BI Gateway comes in two editions: Personal Gateway, designed for individual users or small teams, and Enterprise Gateway, suitable for organizations with multiple users and complex data sources.

**Use Cases:**

* **Hybrid Deployments:** Power BI Gateway facilitates hybrid deployments where organizations have a mix of cloud-based and on-premises data sources. It allows seamless integration of data from both environments.
* **Scheduled Data Refresh:** Power BI Gateway enables scheduled data refresh for reports and dashboards connected to on-premises data sources. This ensures that insights are based on the most recent data.
* **Real-time Analytics:** With Direct Query mode, Power BI Gateway supports real-time analytics by connecting directly to on-premises data sources without data replication.
* **Data Security:** By keeping sensitive data on-premises, Power BI Gateway helps organizations maintain data security and compliance with regulatory requirements.
* **Large-scale Deployments:** Enterprise Gateway is suitable for large-scale deployments in organizations with multiple users and data sources. It provides centralized management and monitoring capabilities for IT administrators.

Overall, Power BI Gateway plays a critical role in enabling organizations to leverage the power of Power BI while securely accessing and refreshing data from on-premises sources. It enhances data connectivity, enables real-time analytics, and ensures data security and compliance.

1. **How would you approach this problem,** **if the objective and subjective questions weren't given?**

In addressing the question if the objective and subjective questions weren't given, will to approach in my way:

1. **Total Crime and Arrest Percentage**: Begin by calculating the total number of crimes and the arrest percentage. This offers an overview of the crime situation and evaluates the effectiveness of law enforcement in addressing these crimes.
2. **Distribution of Crime by Location**: Use a map visualization to understand the distribution of crimes across various locations. This aids in identifying high-crime areas and can inform resource allocation and law enforcement strategies.
3. **Day-Night Crime Percentage**: Analyze the percentage of crimes occurring during the day versus at night. This information is crucial for planning patrols and allocating resources based on the time of day when crimes are most likely to occur.
4. **Analysis by Crime Type and Location**: Further analyze the data by examining the distribution of crime types across different locations. This helps identify patterns and hotspots for specific types of crimes, guiding targeted interventions and prevention efforts.
5. **Calculation of Resolution Time**: Calculate the resolution time for each case by subtracting the date of occurrence from the date updated. This metric provides insight into the efficiency of law enforcement in resolving cases and may highlight areas for improvement.
6. **Analysis of Domestic Crime**: Investigate the proportion of domestic crimes and the factors influencing their resolution. Understanding the dynamics of domestic crimes is essential for developing targeted strategies and interventions.

By employing these analytical approaches, a comprehensive understanding of the crime data can be achieved, enabling more effective and strategic law enforcement responses

1. **If you are also given a table of districts-states with state\_id, district\_id and name, what would be the type of relationship between district of our data and district\_id of new table?**

Solution:

The type of relationship between the district in our crime data table and the district\_id in the new table of districts-states would likely be a "Many-to-One" relationship.

**Many-to-One Relationship**: This type of relationship implies that multiple rows in the crime data table (many districts) can be related to a single row in the districts-states table (one district\_id). In other words, each district\_id in the districts-states table uniquely identifies a district, and multiple records in the crime data table may correspond to the same district\_id.