**Objective Questions**:

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

**Ans: -** To clean the data set First I have loaded the Data Set in the Power BI Application and then have used Power Query Editor option and loaded the data in the same.

The options I have used in power Query Editor are as follows: -

# Clicked to the Manage option in the Home ribbon tab and clicked on Duplicate option in the same.

# After performing the same I then have filtered the null values in Ward column & X Coordinate Column

# After that I have changed the data format of some of the columns named: -

Beat, District, Ward, Community Area, X Coordinate, Y Coordinate, & Year to number format.

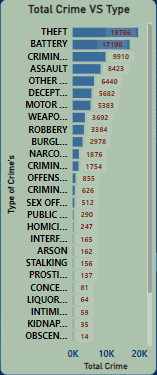
Then I have separated the column named Update on in two columns and have named it as updated on date & updated on time.

I have separated just the date and time in two different columns and renamed it.

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

**Ans: -** For Answering the same question I have: -

* Conducted analysis of crime types by calculating a Total Crime measure
* Visualized the findings using a clustered column chart.
* Identified **Theft** as the most prevalent crime, with 89898 reported incidents.



The answer is provided in the Locality Tab chart named: - Total Crime VS Type.

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

Ans: - To answer this question I have created a measure to calculate the percentage of reported incidents that have resulted in an arrest which has resulted as 12.42% the formula used is: -

Arrest Rate = AVERAGE('crimes\_data\_2022 - crimes\_data\_'[Arrest In numbers])



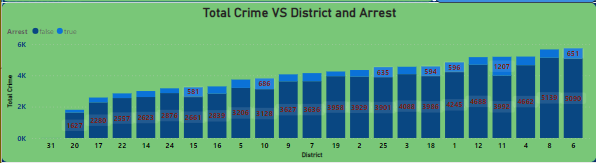
The answer is provided in the Main Tab Arrest Rate

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.**

Ans: - To calculate the number of crimes in each district I have used Stacked column Chart and have seen that the **district 6** is having the highest crime rate as compared to other district.

The following steps are being taken: -

* Utilized a Stacked Column Chart to analyse the number of crimes in each district.
* Identified District 6 as having the highest number of crime incidents.
* Total reported cases in District 6 amounted to 5,059.



The answer is provided in the Exploration Tab and the chart named as: -Total Crime VS District and Arrest.

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

Ans: - Identified multiple categorical columns in the dataset, including:

* Case Number
* Block
* IUCR
* Type
* Location Description
* Arrest
* Domestic
* Beat
* District
* Ward
* Community Area
* FBI Code
* Year

These columns categorize different aspects of the data, offering valuable insights for analysis.

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?

Ans: - Observed 944 null values in both the X-coordinate, Y-coordinate and also in the ward columns.

* These null values account for approximately 1% of the total dataset.
* To preserve data integrity and considering the small proportion, it was decided to remove these null values.
* 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.

Ans: -

* Determined that Domestic Crime comprises 20.02% of the total crimes in Chicago.
* Incorporated a card visual in the 'Main' tab to showcase this insight.
* Utilized the 'Domestic Crime Proportion' metric with the following DAX formula: -
* Crime Prroportion Domestic = DIVIDE(SUM('crimes\_data\_2022 - crimes\_data\_'[Domestic in Numbers]),COUNTROWS('crimes\_data\_2022 - crimes\_data\_'))

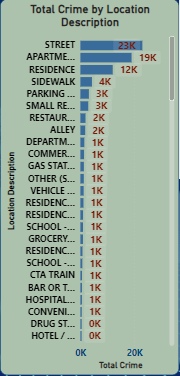


* The visual offers a clear representation of the proportion of domestic crimes within the overall dataset.

The answer is provided in the Main Tab named as: -Crime proportion Domestic.

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.

Ans: - To analyse the distribution of crimes based on location, I utilized a Stacked Bar Chart with "Location Description" on the Y-axis and the total number of crimes on the X-axis. The findings revealed that **streets** had the highest number of crimes compared to other locations.



The answer is provided in the Locality Tab named as: - Total Crime by location Description.

Top of Form

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

Ans: - To determine the time taken to resolve cases, I created a measure and visualized it in the Main Tab. The formula used for this measure is:

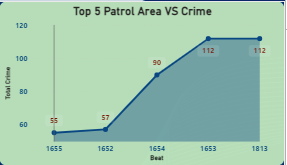
Days taken to resolve = DATEDIFF('crimes\_data\_2022 - crimes\_data\_'[Date],'crimes\_data\_2022 - crimes\_data\_'[Updated On.Date],DAY)



The answer is provided in the Main Tab named as: - Average Days taken to resolve.

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

Ans: - Utilizing an area chart, I analysed the correlation between beats and total reported crimes. Notably, beat 1655 emerged with only 55 reported crimes, indicating a comparatively low incidence of criminal activities in this area, which is managed by patrol officers.



The answer is provided in the Main Tab named as: - Top 5 patrol area VS Crime.

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

Ans: - Created calculated columns and utilized the 'Add Column' feature in Power BI Query.

Generated Longitude and Latitude columns using the "Column from Example" feature, providing sample data separated by a comma as the delimiter.

Determined the **Updated-on** column and have extracted date and time separately.

Developed a total of 9 calculated columns using DAX formulas, incorporating 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.

Key difference between calculated columns and custom columns lies in their application and functionality: -

**Calculated Column: -** Created in Power BI data model using DAX formulas, computed and stored during data loading, available for various calculations and operations within the dataset, and usable in visualizations and analyses.

**Add Column (Custom Column):** - Utilized in Power Query Editor, dynamically generates custom columns based on transformations applied to existing columns during data transformation, not stored in the data model, primarily used for data preparation and enrichment tasks.

In summary, calculated columns and custom columns serve similar purposes, but their creation methods and functionalities differ. Calculated columns are stored in the data model and are computed during data loading, while custom columns are dynamically generated during data transformation in Power Query Editor.

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

Ans: - The calculation for Theft in District 8, reported at 1143 incidents, is derived using the following formula:

Theft Crime in Dis 8 = CALCULATE([Total Crime],AND('crimes\_data\_2022 - crimes\_data\_'[Type]="Theft",'crimes\_data\_2022 - crimes\_data\_'[District]="8"))



This formula specifically targets 'Theft' type crimes within District 8, resulting in a precise count of 1143 incidents.

The answer is provided in the Main Tab named as: - Theft Crime in Dis 8.

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

Ans: - I used the 'Add Column with Example' feature in Power Query Editor to split values using a comma delimiter. This streamlined the process, enabling me to efficiently manipulate and organize the data, ultimately enhancing its suitability for analysis and visualization.

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?

Ans: - When using the 'Add Column' feature in Power Query, the M language code shown in the formula bar depends on the specific transformations applied to the data. Let's consider two examples:

# Splitting Location Column: The M query for splitting the "Location" column into separate columns based on the comma delimiter might appear as follows:

= Table.SplitColumn(#"Changed Type1", "Location.2", Splitter.SplitTextByDelimiter("-", QuoteStyle.None), {"Location.2.1", "Location.2.2"})

* The Table.SplitColumn function is utilized to split the "Location" column into two separate columns, "Location.1" and "Location.2", based on the comma delimiter.
* "#Filtered Rows" refers to the preceding step in the query where rows may have been filtered.

Top of Form

The M-query, also referred to as the Power Query Formula Language, serves as a functional language utilized for data transformation in Power Query. It empowers users to define sequential transformations on their data within the Power Query Editor. M-query encompasses a diverse set of operations including data type conversions, column splitting, table merging, row filtering, and beyond.

With M-query, users have the capability to execute intricate data transformations and perform data cleansing tasks to refine their data for subsequent analysis. This language offers a robust and adaptable framework for manipulating data within Power Query, allowing users to devise custom data transformation workflows tailored to their specific needs.

**Subjective questions**

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

Ans:- Upon analyzing the month-wise change in crime rates from January to June 2022, a distinct pattern emerges. The data reveals a gradual increase in crime rates leading up to May, followed by a significant decrease in June. It's important to note that the sharp decline in June could be attributed to the limited data available, only up to the 9th of June.

Specifically, in May, the recorded crime rate peaked at 19,353 incidents, indicating a month with a higher-than-average crime rate during the observed period. This finding underscores the importance of investigating the factors contributing to this surge in criminal activities during May.

Additionally, the analysis highlights that February experienced a higher arrest rate of 13.83%. While this may suggest improved law enforcement efficiency, it's essential to delve into the dynamics behind this peak in arrests during February.

Focusing on addressing the contributing factors to the surge observed in May could serve as a pivotal strategy in reducing overall crime in Chicago.



The answer is provided in the Main Tab named as: - Total Crime in a Month.

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?

Ans: - Analysis reveals theft and battery as the most prevalent crime types, with theft accounting for 19,776 incidents and battery for 17,190.

Arrest rates show a notable contrast between the two categories—4.08% for theft and 13.92% for battery.

Domestic crime plays a significant role, constituting 5.40% of theft incidents and a notably higher 54.64% of battery incidents.

Strategic areas for crime reduction efforts:

**# Policing Initiatives:** Strengthen and optimize policing in areas with high theft and battery incidents.

**# Surveillance and Technology:**

Increase surveillance in high-theft areas using CCTV cameras and smart city initiatives.

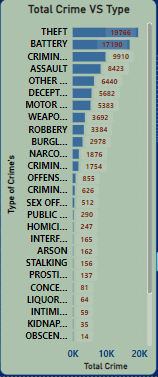
Implement advanced analytics for predictive policing.

Legal Enforcement:

Review and enhance legal enforcement measures for a swift response to criminal activities.

Aim to increase arrest rates as a deterrent for potential offenders.

**Conclusion:** Crime reduction efforts require a multi-faceted approach, focusing on high-incidence areas, domestic crime prevention, and reinforcing legal and policing frameworks for a safer Chicago.



The answer is provided in the Locality Tab named as: - Total Crime VS Type.

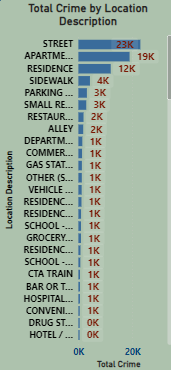
1. Which localities experience higher crime rates, and what measures can we ensure to reduce these numbers?

Ans: - Analysis of crime data underscores specific localities, such as Apartment Street and Residence Location, exhibiting higher crime rates, with incidents totaling 14,768, 13,910, and 8,345, respectively. These areas notably witness a prevalence of theft and battery crimes, accompanied by suboptimal arrest rates. The analysis also hints at potential societal influences on arrest rates, particularly evident in domestic crime cases.

To mitigate and diminish crime in these localities, several targeted measures can be implemented:

* **Enhanced Surveillance:**
  + Increase police presence and surveillance, especially in high-crime areas like Apartment Street and Residence Location.
  + Employ advanced surveillance technology, such as CCTV cameras, to monitor and deter criminal activities.
* **Infrastructure Improvement:**
  + Enhance street lighting to augment visibility during nighttime, thereby reducing opportunities for criminal activities.
  + Focus on urban planning strategies, such as well-lit public spaces, to deter crime.
* **Addressing Root Causes:**
  + Invest in education and job training programs to provide economic opportunities, addressing underlying causes of crime.
  + Ensure the availability of accessible mental health and substance abuse services to support individuals in need.
* **Youth Engagement Programs:**
  + Implement proactive youth engagement initiatives to steer young individuals away from criminal activities.
  + Advocate for and implement effective gun control measures to prevent unauthorized access to firearms.

By embracing a comprehensive, community-oriented approach that integrates law enforcement strategies with social and economic interventions, these localities can witness a significant reduction in crime rates. Addressing both immediate contributing factors and underlying root causes is essential for effecting lasting positive change in these areas.



The answer is provided in the Locality Tab named as: - Total Crime by Location Description.

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

Ans: - Analysis from the stacked column chart reveals that Wards 42, 27, 28, 6, and 24 exhibit elevated crime rates, each surpassing 3000 incidents. Additionally, Ward 7 stands out with a notably high percentage of domestic crime. To enhance security and diminish crime in these wards, the following measures can be implemented:

**#Enhanced Police Presence:**

* + Deploy additional law enforcement officers and resources to actively patrol and deter criminal activities in the identified high-crime wards.

**# CCTV Surveillance:**

* + Strategically install surveillance cameras to monitor and record activities, acting as a deterrent and aiding in investigations.

**#Mental Health and Substance Abuse Services:**

* + Ensure accessibility to mental health and substance abuse support services to address underlying issues contributing to criminal behavior.

**#Targeted Security Measures in Specific Areas:**

* + Implement tailored security measures in various areas such as public transportation hubs, residential neighborhoods, commercial districts, parks and recreational areas, schools and educational institutions, and high-crime zones.

**#Cybersecurity**:

* + Address the modern threat landscape by investing in cybersecurity infrastructure, raising awareness about online safety, and providing resources for cybercrime prevention.

By implementing this comprehensive set of measures, a more effective and nuanced approach can be taken to reduce crime rates in the identified wards, fostering a safer and more secure environment for residents.

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

Ans: - Analysis of crime data from 2022 reveals fluctuations in crime rates throughout the year, with May exhibiting a significant peak.

* Weekly trend analysis identifies Tuesdays and Saturdays as peak days for criminal activities.
* While these insights are valuable, it's crucial to recognize the limitations of analyzing data from a single year.
* For a more accurate understanding of discernible patterns or trends, it's advisable to gather and analyze data spanning multiple years.
* Longitudinal data facilitates a comprehensive examination of fluctuations, seasonal variations, and long-term trends in crime rates.
* Crime appears to be evenly distributed across various areas, making it challenging to discern specific trends or patterns.
* However, a notable disparity emerges when analyzing crime occurrence during different times of the day.
* Crime rates notably increase during nighttime hours, signaling a need for proactive measures during this period.
* Addressing nighttime crime necessitates proactive efforts from authorities, such as increased patrolling, improved lighting in vulnerable areas, and tailored community engagement initiatives focused on nocturnal activities.
* By directing resources and attention toward nighttime crime prevention strategies, authorities can effectively reduce overall crime rates in the city and enhance public safety.

1. Create a monthwise 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)?

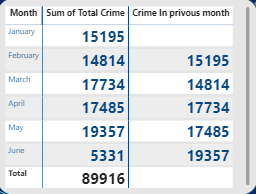
Ans: - In my Power BI project, I encountered a scenario where I needed to analyze crime data on a month-by-month basis. To achieve this, I initially created a continuous date table to ensure effective utilization of DAX functions. However, I realized that I needed a different approach to calculate the total number of crimes for the previous month.

Subsequently, I created a new table named "Dates\_crime," which contains all distinct dates associated with the crimes that occurred on those specific dates. I then 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 visually represent the relationship between the current and previous months' crime data, I utilized a table in Power BI. This table facilitated the comparison of 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 effectively utilizing DAX functions, I gained valuable insights into month-to-month variations in crime occurrences. These insights are crucial for making informed decisions and implementing targeted interventions.

This is the DAX formula used: - Crime In privous month = CALCULATE(SUM(Crime\_Date[Total Crime]),PREVIOUSMONTH('date'[Date]))



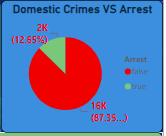
The answer is provided in the Locality Tab named as: - Crime in previous month.

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?

Ans: - The data analysis underscores a concerning pattern regarding domestic crimes and arrest rates. According to the pie chart visualization, only 12.65% of domestic crimes result in arrest. This finding aligns with previous observations indicating that many domestic crimes go unreported and subsequently lack arrests due to factors such as public reluctance and familial pressure.

The low arrest rate for domestic crimes highlights a significant gap in addressing this specific category of offenses. Factors like public hesitation to report incidents and familial pressures contribute to underreporting and the subsequent lack of arrests.

To effectively address this trend, it's crucial to implement measures aimed at encouraging the reporting of domestic crimes, providing support for victims, and ensuring that law enforcement agencies are adequately trained and equipped to handle such cases sensitively and efficiently. Additionally, raising awareness about the importance of reporting domestic incidents and breaking the cycle of silence within families is essential for improving arrest rates and ensuring justice for victims of domestic violence and abuse.

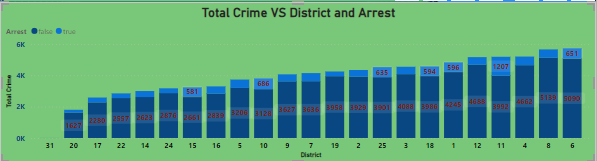


The answer is provided in the Exploration Tab named as: - Domestic Crimes VS Arrest.

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

Ans: - Examination of the description column reveals several frequently occurring terms:

* + 'Simple' with 11,244 instances
  + 'Domestic Battery Simple' with 8,060 occurrences
  + '$500 And Under' with 7,024 instances
  + 'Over $500' with 6,813 appearances
* This visual representation effectively highlights these common terms, providing insights into prevalent incident types.
* Stakeholders can utilize this visualization to gain detailed understandings of the distribution of these frequently encountered terms.
* It's important to note that the visualization includes a filter option, allowing users to concentrate on specific categories.
* Removing the filter broadens the scope, offering a comprehensive depiction of term frequencies across all categories.



The answer is provided in the Exploration Tab named as: - Total Crime VS District and Arrest.

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

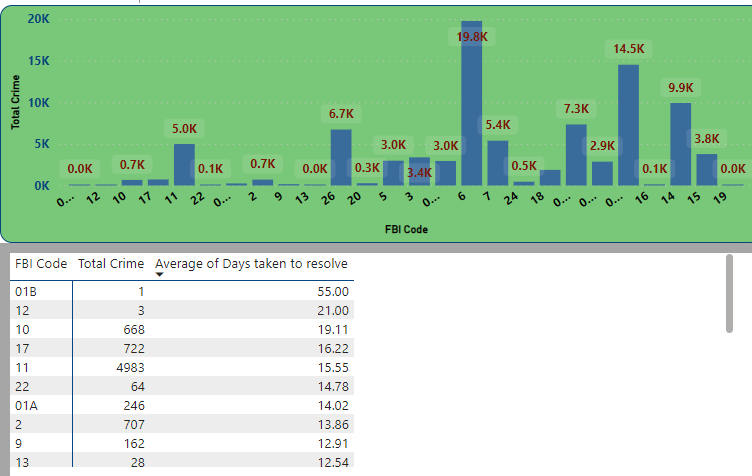
Ans: - The dataset does not specify any particular regions or geographic areas as being explicitly "domestic." Domestic crimes are typically categorized based on the type of crime (such as domestic violence or disturbance) rather than specific locations. Therefore, solely relying on the dataset, it's not feasible to pinpoint regions with significantly high reported domestic crime rates.

Identifying areas with elevated rates of domestic crime would necessitate a more in-depth analysis that considers location descriptions, addresses, or community areas where such incidents occurred. This analysis would involve filtering and categorizing incidents related to domestic crimes and then examining their prevalence in specific geographical areas.

It's crucial to recognize that domestic crimes can occur across various neighborhoods and communities. Addressing this issue effectively often requires collaborative efforts between law enforcement, social services, and community organizations to offer support and intervention for victims while ensuring accountability for offenders.

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

Ans: -



* The visual representation illustrates the relationship between crime type, locality, FBI codes, and average case resolution time.
* Crimes such as license violations and gambling are resolved more quickly, while human trafficking and obscenity cases typically take longer to resolve.
* Incidents at airport terminals and on boats or watercraft are resolved relatively swiftly compared to cases associated with pawn shops and colleges.
* FBI codes 19, 15, and 14 are notable for efficient case resolution, indicating effective handling of crimes categorized under these codes.
* Understanding the specific nature of each crime, as categorized by FBI codes, is crucial in determining resolution time.
* Prioritizing efficient strategies for addressing crimes associated with these codes can improve overall case-solving efficiency for law enforcement agencies.

The answer is provided in the Exploration Tab named as: - Total Crime and Average of Days taken to resolve by FBI Codes.

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

Ans: - I've successfully implemented a calculated column using the formula:

Question11 = CALCULATE(COUNTROWS('crimes\_data\_2022 - crimes\_data\_'),FILTER('crimes\_data\_2022 - crimes\_data\_','crimes\_data\_2022 - crimes\_data\_'[Type]="Theft"&&'crimes\_data\_2022 - crimes\_data\_'[District]=8))

This calculated column is instrumental in categorizing data for further analysis, facilitating the identification and examination of domestic incidents specifically within District 8.

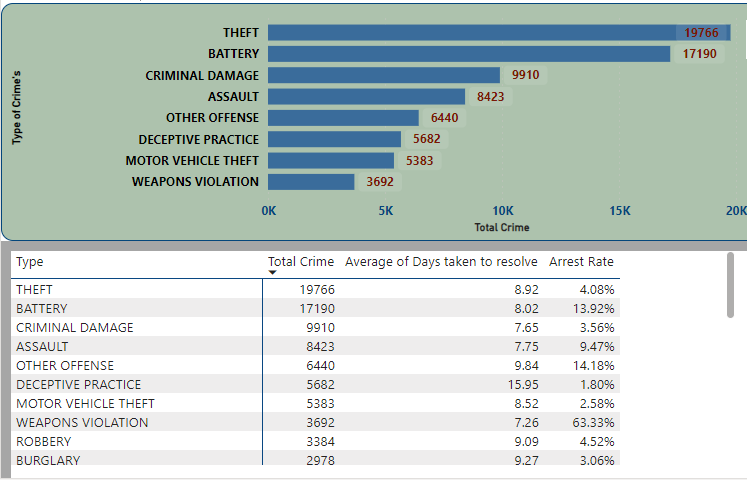
1. Out of all the types of crimes which do you think is the most dangerous one and rank the type of crimes according to their no. of occurrences?

Ans: - Upon scrutinizing the dataset, it's evident that theft, battery, and criminal damage are among the most frequently occurring crimes. Despite their high prevalence, these offenses also exhibit notably low rates of arrest.

Among these crimes, theft stands out as particularly concerning due to its frequency and potential impact on individuals and communities. Not only does it inflict financial losses on victims, but it also fosters feelings of insecurity and erodes trust within society.

When ranking the types of crimes based on their occurrence rates, theft would likely claim the top spot, followed by battery and criminal damage. This ranking underscores the severity of these offenses in terms of both their prevalence and societal ramifications.

Furthermore, the observed low arrest rates for these crimes highlight the challenges encountered in effectively addressing and combating them.



The answer is provided in the Locality Tab named as: - Total Crime VS Type.

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

Ans: - The Power BI Gateway, offered by Microsoft, acts as a bridge connecting on-premises data sources with the cloud-based Power BI service. It facilitates secure access and data refresh from on-premises sources to Power BI.

Key Aspects:

* Data Connectivity: Power BI Gateway enables connection to various on-premises data sources like SQL Server, Oracle, and SharePoint, establishing a secure link with the Power BI service in the cloud.
* Data Refresh: It supports scheduled data refresh for on-premises sources, ensuring Power BI reports and dashboards remain updated with current data.
* Direct Query: Power BI Gateway supports real-time querying of on-premises data sources, enabling analysis without data replication.
* Personal and Enterprise Editions: It comes in two editions catering to individual users/small teams (Personal Gateway) and organizations with multiple users and complex data sources (Enterprise Gateway).

Use Cases:

* Hybrid Deployments: Facilitates integration of both cloud-based and on-premises data sources.
* Scheduled Data Refresh: Ensures reports and dashboards are regularly updated with the latest data.
* Real-time Analytics: Supports real-time analysis by querying on-premises data sources directly.
* Data Security: Helps maintain data security and regulatory compliance by keeping sensitive data on-premises.
* Large-scale Deployments: Enterprise Gateway provides centralized management and monitoring for organizations with multiple users and data sources.

In summary, Power BI Gateway is essential for organizations leveraging Power BI, enabling secure access and data refresh from on-premises sources, enhancing data connectivity, and ensuring compliance

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

Ans: - I would have followed the given below approaches: -

1. Total Crime and Arrest Rate: Begin by computing the overall number of crimes and the percentage of arrests. This offers an overview of the crime landscape and assesses law enforcement effectiveness in addressing these incidents.
2. Crime Distribution by Location: Employ a map visualization to comprehend the spread of crimes across various locations. This aids in identifying high-crime areas, informing resource allocation, and shaping law enforcement strategies.
3. Day vs. Night Crime Percentage: Evaluate the proportion of crimes occurring during daytime versus nighttime. This data informs patrol scheduling and resource deployment based on peak crime hours.
4. Crime Type and Location Analysis: Delve deeper into the dataset by examining how different crime types are distributed across locations. Identifying trends and hotspots for specific crimes assists in targeted interventions and preventive measures.
5. Resolution Time Calculation: Determine the time taken to resolve each case by subtracting the occurrence date from the update date. This metric sheds light on law enforcement efficiency in case resolution, pinpointing areas for potential improvement.
6. Domestic Crime Analysis: Investigate the prevalence of domestic crimes and the factors affecting their resolution. Understanding the dynamics, including familial pressures, is essential for devising effective support and intervention strategies.
7. 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?

Ans: - Including the district\_id in the new table would establish a many-to-one relationship between the current table and the new one. In the dataset, each district\_id recurs multiple times as each district encompasses numerous associated cases. Hence, each district in the data corresponds to only one district\_id in the new table, thereby forming a many-to-one relationship.