**BY MALLESH H**

Objective questions

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

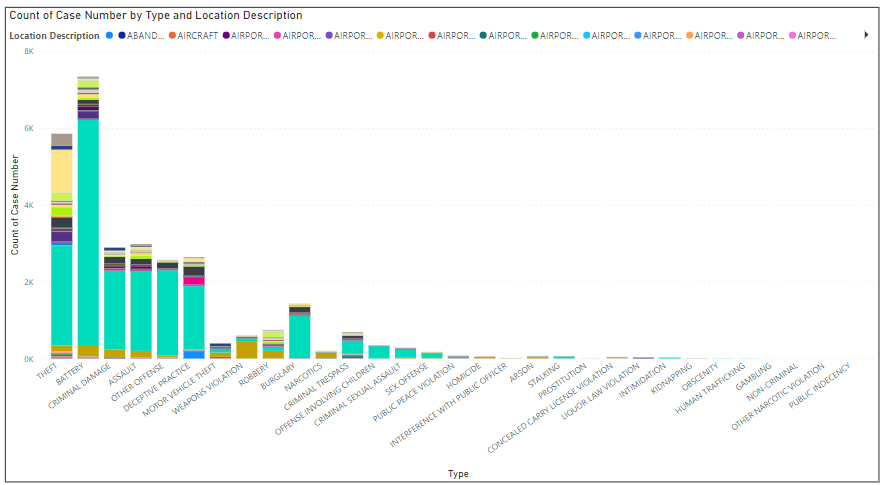
I removed the X coordinate and Y coordinate columns from the data set because they are not necessary for the analysis or interpretation of the data. These columns do not contribute to the insights we seek to gain from the data set, so they were excluded to streamline the data and focus on relevant information.

Null values were present in the ward column, and blank values existed in the location column within the given data set. These discrepancies were addressed by utilizing the filter option during the transformation process prior to loading the data.

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

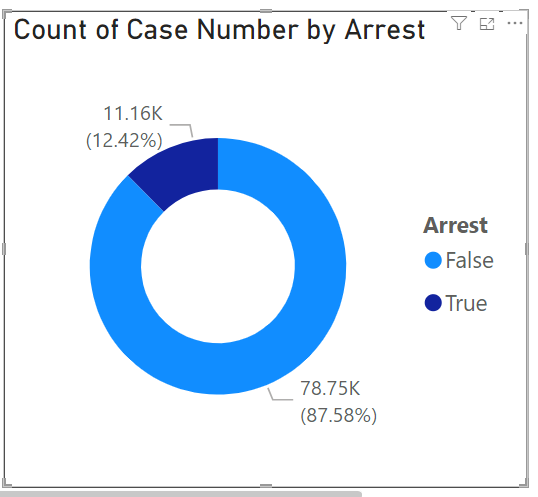
For crime type analysis, I utilized a clustered column chart where the crime types were plotted along the x-axis. The count of case numbers was represented along the y-axis, and the location descriptions were depicted in the chart legend. This visualization enables a clear comparison of case counts for different crime types across various locations.

Visualization chart 🡪 Combined column chart



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

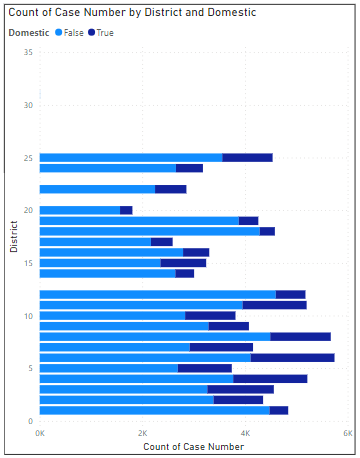
For the evaluation of arrest rates, I created a donut chart where the count of cases was plotted in the values section, and the "arrested" column was represented in the legend. This visualization allows us to analyze the percentage of reported incidents that have resulted in an arrest, providing insights into law enforcement effectiveness. by analyzing the donut chart, it became evident that the percentage of arrest rates was considerably lower compared to the total number of cases registered.

Visualization chart 🡪 Donutchart c

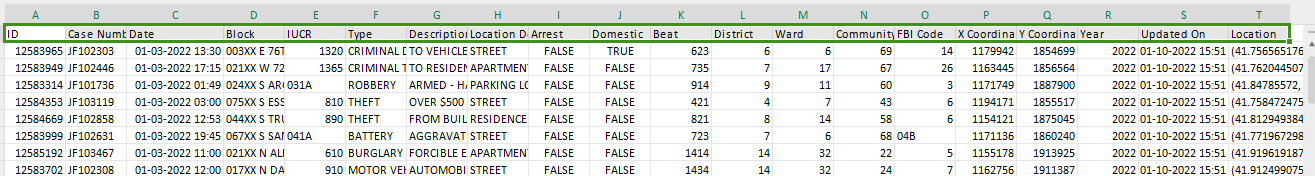
1. District Crime Distribution Assessment: Calculate the number of crimes in each district to understand how crime is distributed across the city and identify high-crime areas.

I created a bar chart to analyze the distribution of the number of cases across each district. In this visualization, the number of cases was plotted along the x-axis, while the district numbers were represented along the y-axis. This visualization allows for a clear understanding of the distribution of cases among different districts, enabling insights into the varying levels of reported incidents across the area.

Visualization chart 🡪 Bar chart

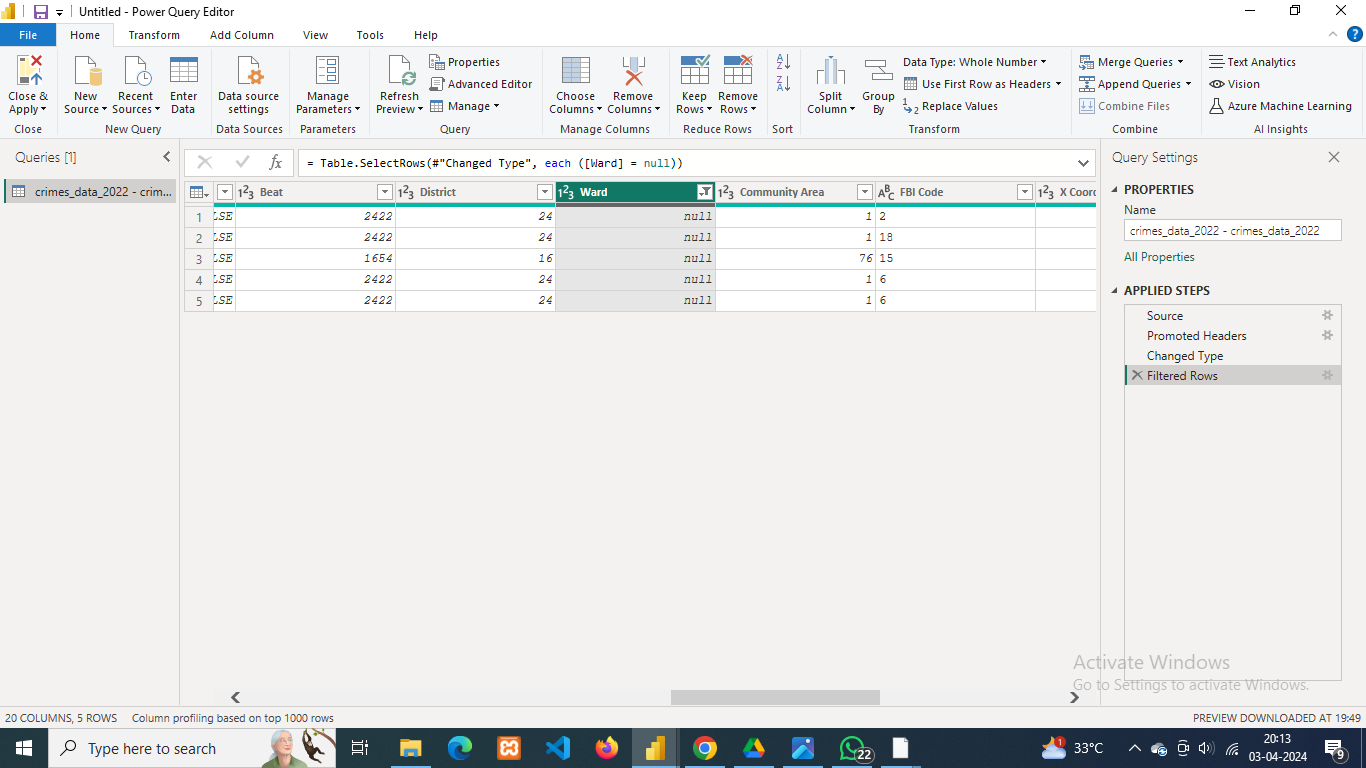


1. How many categorical attributes are there in the data?

In the given data set, there are 20 categorical attributes initially.

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?

Null values were identified in the ward column of the given dataset. To address this issue, I filtered out the null values in the ward column during the transformation process before loading the data.



Handling null values in Power BI depends on the specific context and requirements of your analysis. However, here are some common approaches to handling null values in Power BI:

**Replace with a Default Value:** Substitute null values with a default value that is meaningful in the context of your analysis. For example, you might replace null values in numerical columns with zero or null values in text columns with "Unknown".

**Remove Null Values:** If null values do not provide any useful information for your analysis, you can choose to remove rows containing null values from your data set. Power BI provides options to filter out or remove rows with null values using filters or data cleaning operations.

**Impute Missing Values:** Use statistical methods to impute missing values based on the existing data. This could involve filling null values with the mean, median, or mode of the respective column. Power BI offers various functions and tools for data transformation and imputation.

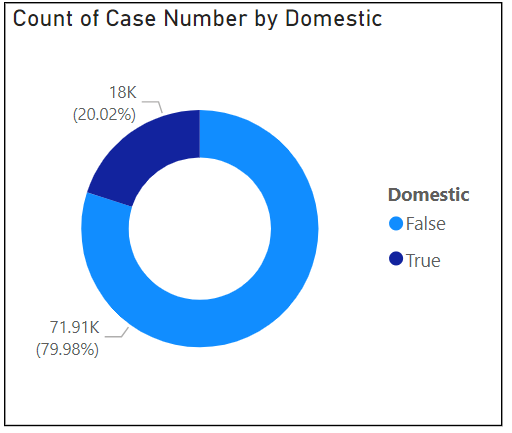
**Flag Null Values:** Create a new column to flag null values in the dataset for further analysis or visualization. This approach allows you to retain the information about null values while also highlighting their presence in the data.

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

I utilized a donut chart to analyze the proportions of domestic and non-domestic crimes. In this visualization, I plotted the number of cases in the values field, with the domestic column represented in the legend. This allowed for a clear comparison of the proportions of domestic and non-domestic crimes within the dataset.

By analyzing the donut chart, it reveals that domestic crimes constitute 20% of the total, whereas non-domestic crimes account for 80%. This indicates a ratio of domestic to non-domestic crimes of 1: 4.

Visualization chart 🡪 Donut chart

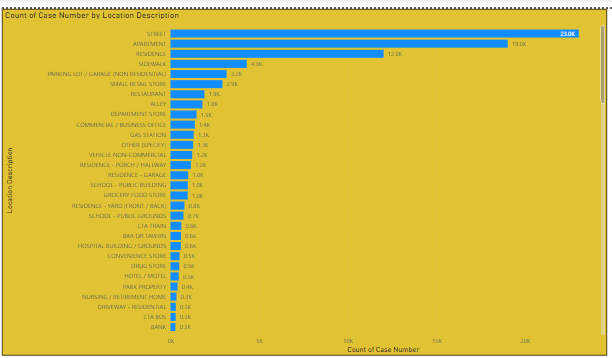


1. Is there any “Location Description” where the number of crimes is higher than expected? Come up with a table or visualization in which one can judge the frequency of crimes at each Location Description type.

I created a bar chart to analyze the distribution of the number of cases across different areas. In this visualization, the number of cases was plotted along the x-axis, while the location descriptions were represented along the y-axis. This allowed for a clear understanding of how the number of cases is distributed across various areas.

The number of crimes reported in apartments was unexpectedly high, totaling 19,019 cases, making it the second most common location description compared to others. This observation is surprising given that apartments typically have heightened security measures in place.

Visualization plane 🡪 Bar chart



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

First, I calculated the number of days taken to solve each case by creating acalculated column using the DATEDIFF function between the 'Date' column and the 'Update on' column. I named this calculated column 'case duration days'.

Afterwards, I created a new measure by applying the AVERAGE formula to the 'case duration days' column in order to find the average time to solve a case.

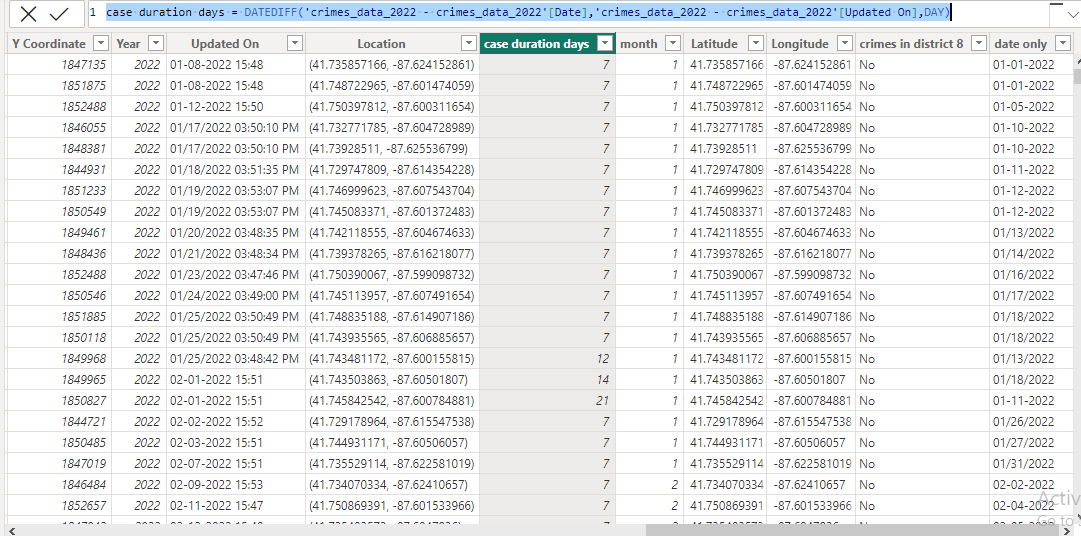
Formulas

To find number of days to solve case:

case duration days = DATEDIFF('crimes\_data\_2022 - crimes\_data\_2022'[Date]'crimes\_data\_2022 - crimes\_data\_2022'[Updated On],DAY)

To find average numbe of days to solve case:

Average days for solving case = AVERAGE('crimes\_data\_2022 - crimes\_data\_2022'[case duration days])

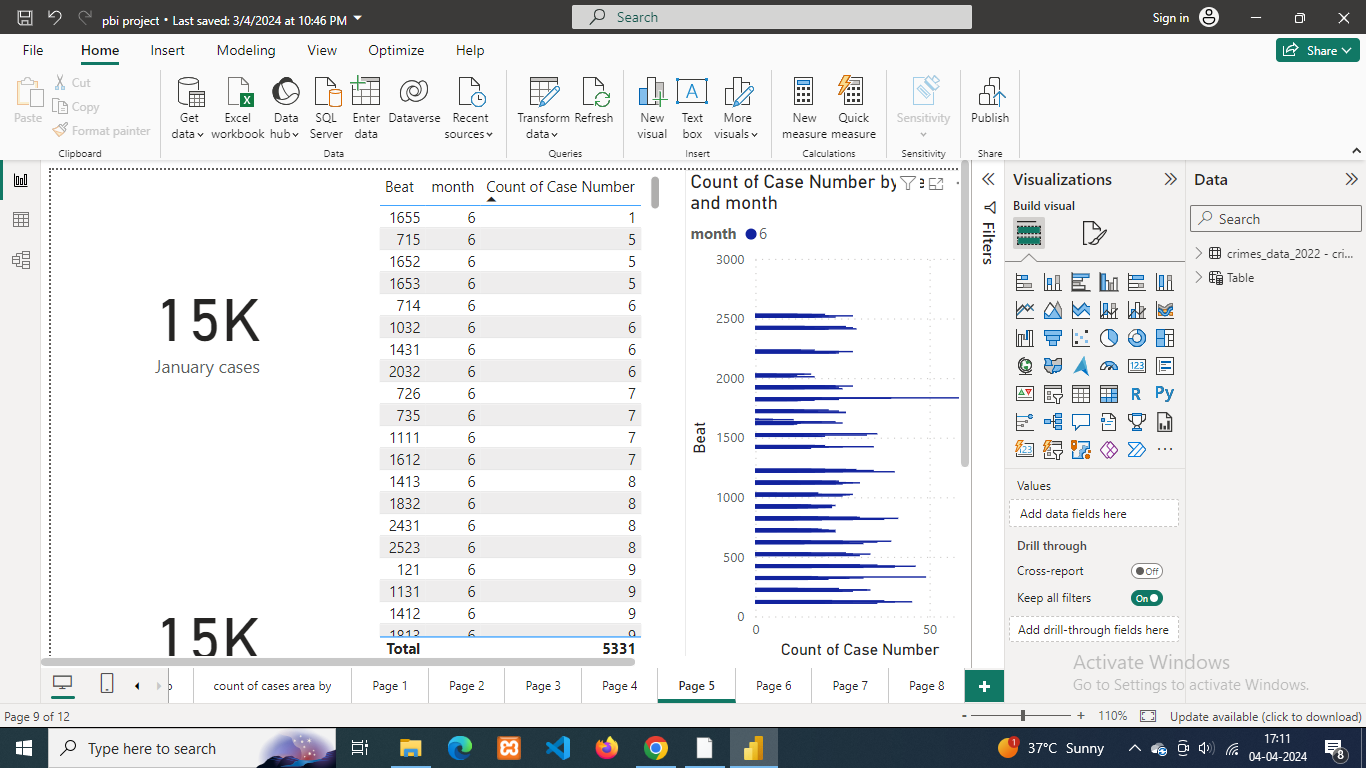


1. To reward the patrol officers, find the patrol area where the crimes reported were under control.

The patrol areas under control are represented by beat numbers 715, 1652, 1653, 1654, and 1655. These areas are identified based on having five or less than five cases reported in the last month according to the provided data.

To find the patrol areas under controll, I utilized matrix table in wich month field applied in the column, beat field in row and count of cases in values also sorted the count of case number in descending order to get the least cases numbers registred beat no.

Visualization chart 🡪 Matrix



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

I have created two calculated columns one is ‘case duration days’ column, another one is ‘month’ column.

**Add column**

* "Add Column" is a feature in Power BI that allows you to add a new column to your data table based on values from existing columns.
* You can use "Add Column" to perform various data transformation tasks such as extracting sub strings from text, converting data types, creating conditional columns, merging columns, etc.
* The operations performed with "Add Column" are applied row by row, based on the values in each row of the existing columns.
* The new column added using "Add Column" is static and its values are not recalculated when the underlying data changes.

**Calculated column**

* A calculated column is a column in a table that is created based on a calculation formula or expression.
* Unlike "Add Column", calculated columns are created using Data Analysis Expressions (DAX), a formula language used in Power BI.
* Calculated columns are evaluated row by row, and the calculation is performed for each row individually based on the DAX expression provided.
* Calculated columns can be used to derive new values, perform mathematical calculations, concatenate strings, apply conditional logic, etc.
* Calculated columns are dynamic, meaning that their values are recalculated whenever the underlying data changes or when the report is refreshed.

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

Number of ‘theft’ crimes happened in ‘District 8’ is 1143.

Formula 🡪

Thefts in District 8 = CALUCULATE (COUNT (‘crimes\_data\_2022 –

crimes\_data\_2022’ [Case Number]), crimes\_data\_2022 – crimes\_data\_

2022’[District] = 8, ‘crimes\_data\_2022 – crimes\_data\_2022’ [Type] =

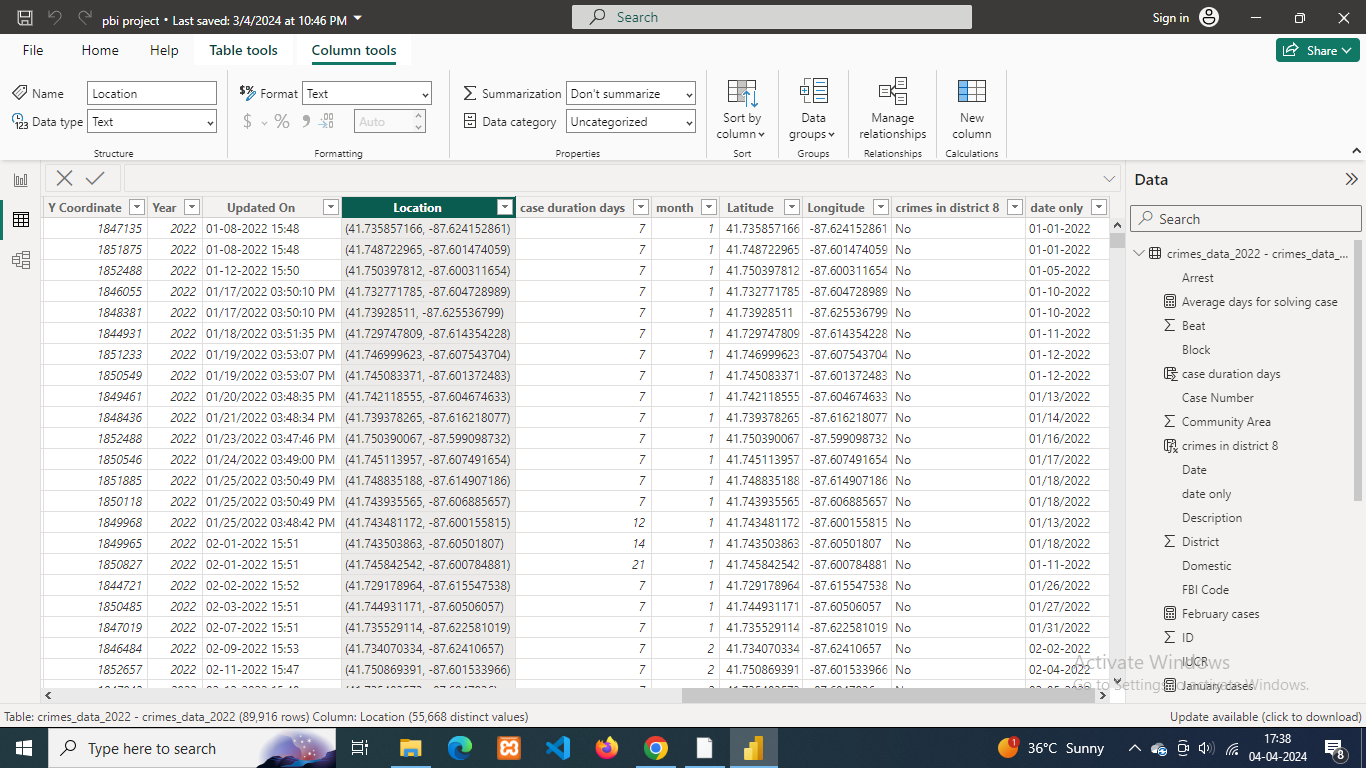
“THEFT”, ALLEXCEPT (‘crimes\_data\_2022 – crimes\_data\_2022’,

‘crimes\_data\_2022 – crimes\_data\_2022’ [Case Number]))

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

Yes, you can separate latitude and longitude from a location column using the split column function.

To retain the original location column unchanged, I duplicated it. Then, I split the duplicate column into two separate columns using a delimiter (-). I named the two resulting columns "latitude" and "longitude."



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?

In Power Query, when you add a column using the user interface, the corresponding M language code appears in the formula bar. The M language, also known as the Power Query Formula Language, is a functional language used to manipulate and transform data within the Power Query Editor.

For example, if you were to add a custom column in Power Query to concatenate two existing columns, the M language code might look something like this:

= Table.AddColumn(#"PreviousStep", "NewColumn", each [Column1] & [Column2])Top of Form

Here, Table.AddColumn is a function that adds a new column to a table. #"PreviousStep" refers to the previous step in the data transformation process, and "NewColumn" is the name of the newly created column. each [Column1] & [Column2] is the expression that defines the logic for the new column, which in this case concatenates the values from two existing columns.

M-query, or simply M, is a powerful language used within Power Query to perform data transformation tasks such as filtering, sorting, grouping, and merging. It provides a wide range of functions and operators for manipulating data structures like tables, lists, and records. M-query is primarily used in Power Query within tools like Microsoft Excel, Power BI, and other Microsoft products for data preparation and transformation before analysis and reporting.

In summary, M-query is a functional language utilized within Power Query to manipulate and transform data, and when you add a column or perform any other data transformation action within Power Query, the corresponding M language code is generated and visible in the formula bar.

窗体顶端

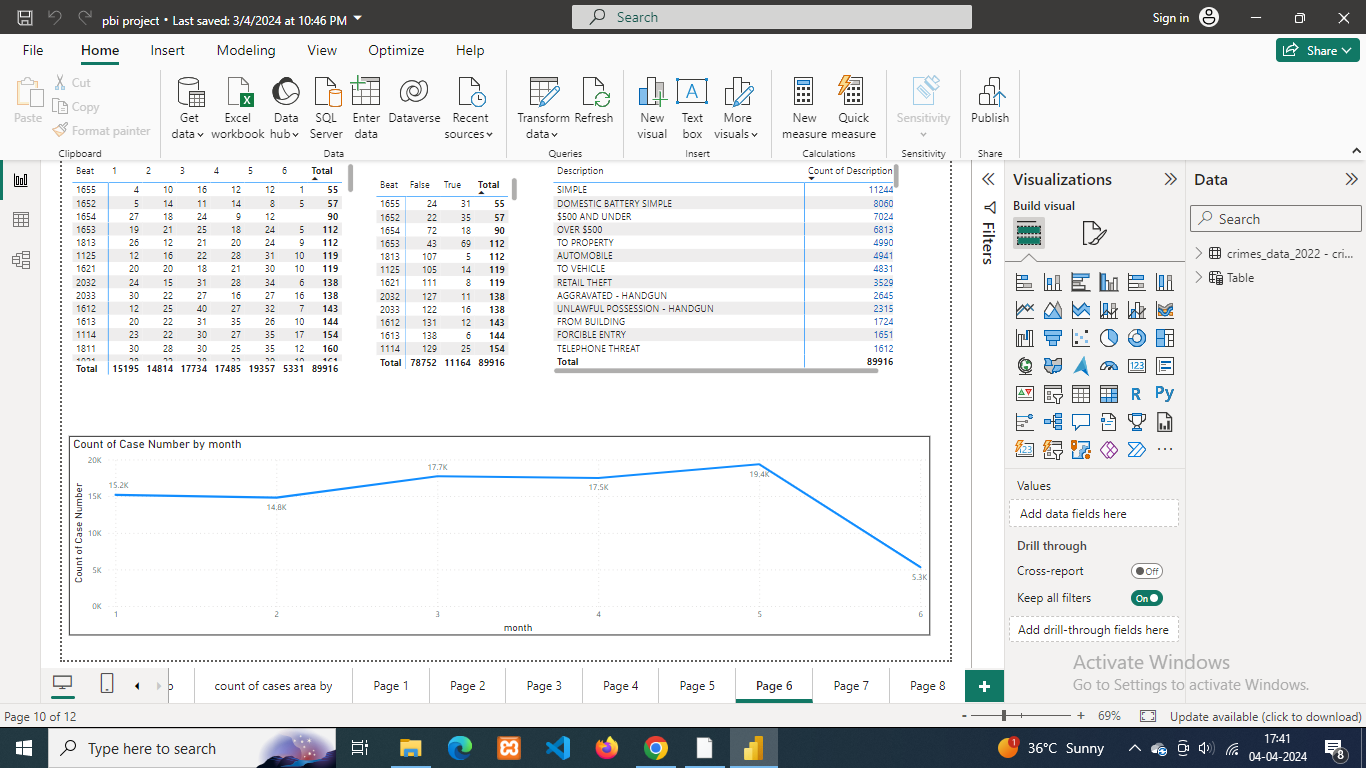
窗体底端

Subjective Questions

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

Yes, there is a month-wise change in crime rates. After analyzing the line chart, it is evident that there are fluctuations in the number of cases reported each month, as depicted by the count of case numbers on the y-axis and the respective months on the x-axis. This indicates that the frequency of reported crimes varies from month to month, suggesting a dynamic pattern in crime rates over time.

Visualisaton chart 🡪 Line chart



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?

To reduce the number of crimes and achieve improvement in overall crime rates, several strategies can be considered:

**Community Engagement and Policing**: Encourage community involvement in crime prevention efforts through neighborhood watch programs, community policing initiatives, and collaboration with law enforcement agencies. Building trust between law enforcement and the community can lead to better crime reporting and cooperation in crime prevention efforts.

**Targeted Enforcement**: Prioritize enforcement efforts in high-crime areas and focus on addressing specific types of crimes that have a significant impact on public safety. Deploy resources strategically to target known crime hotspots and repeat offenders.

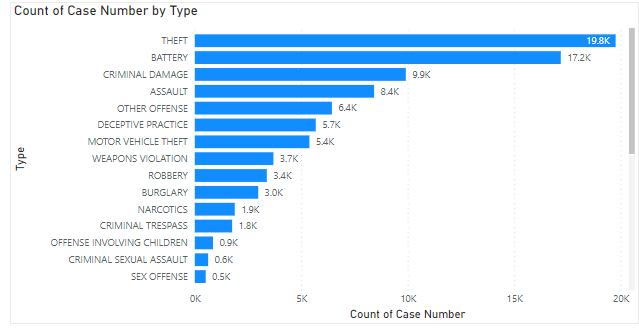
**Crime Prevention Through Environmental Design (CPTED):** Implement urban planning and design principles that discourage criminal behavior and promote safety in public spaces. This includes improving lighting, enhancing natural surveillance, and reducing opportunities for crime through environmental modifications.

**Addressing Root Causes:** Identify and address underlying social, economic, and environmental factors that contribute to crime, such as poverty, unemployment, substance abuse, and lack of access to education and social services. Investing in programs that address these root causes can help prevent crime in the long term.

**Early Intervention and Youth Programs:** Implement early intervention programs and youth initiatives aimed at diverting at-risk individuals away from criminal behavior. Provide educational and recreational opportunities, mentorship programs, and support services to help young people make positive choices and avoid involvement in crime.

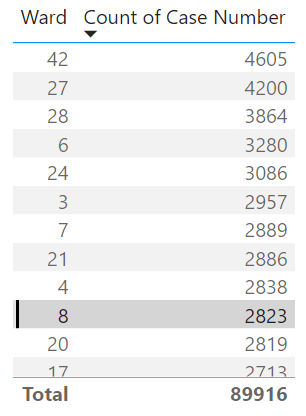
I created a bar chart by plotting type crime in y-axis and count of case numbers in x-axis by analysing the bar chart the type of cases we should consider are theft, battery, criminal damage, and assault type of cases because these are the top 4 case types rigisterded more cases. also we should consider the weapon violation type cases beacuse thise type of cases are one of the most dangerous crimes.

Visualization chart 🡪 Bar chart



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

I created a table using the "ward" and "case number" columns, summarizing the "case" column as a count to distribute the number of cases registered in each ward. The table is sorted by the count of cases to display the wards with the highest number of cases registered at the top. This arrangement allows for easy identification of the wards with the most cases registered.



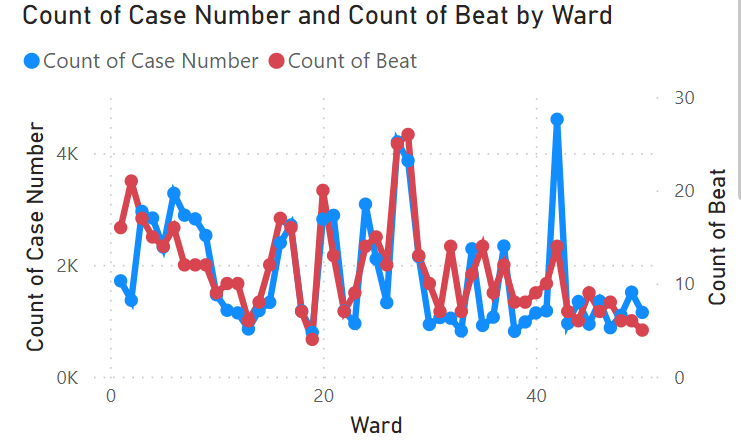
**Increase Police Presence**: Deploy additional police patrols and law enforcement resources to high-crime areas to deter criminal activity and improve response times to incidents.

**Enhanced Surveillance and Technology**: Utilize surveillance cameras, license plate readers, and other technology-driven solutions to monitor high-crime areas and gather evidence for investigations. Implement crime mapping and predictive analytics tools to identify patterns and allocate resources effectively.

**Improve Lighting and Environmental Design**: Enhance lighting and visibility in public spaces to deter criminal activity and increase feelings of safety among residents. Implement crime prevention through environmental design (CPTED) principles to make neighborhoods less conducive to crime and more welcoming to residents.

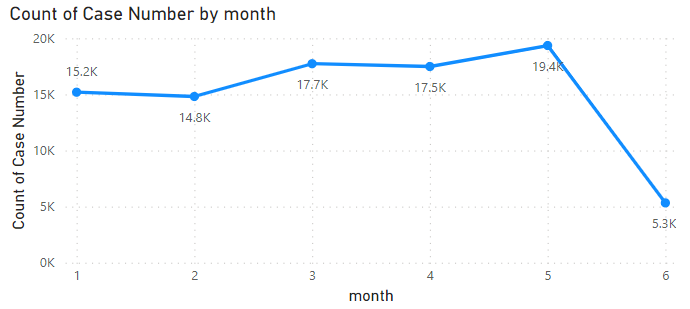
1. Can you suggest wards where security improvements should be made to reduce crime?

To analyze which wards require more security, I created a line chart plotting the count of case numbers and count of beats on the y-axis, with wards represented on the x-axis. From this analysis, it was found that wards with numbers 6, 7, 8, 42, 48, and 50 require additional security measures.

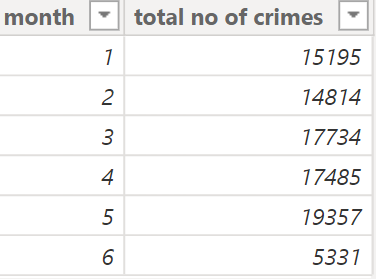


1. Crime Rate Trend Analysis: Monitor changes in crime rates over time to detect any discernible patterns or trends.

To conduct crime rate trend analysis, a line chart was utilized with the month column plotted on the x-axis and the case number column on the y-axis. The case numbers were summarized to count the cases for each month. This approach enabled a visual representation of the trends in crime rates over time, allowing for a clear analysis of how the number of cases fluctuated from month to month.



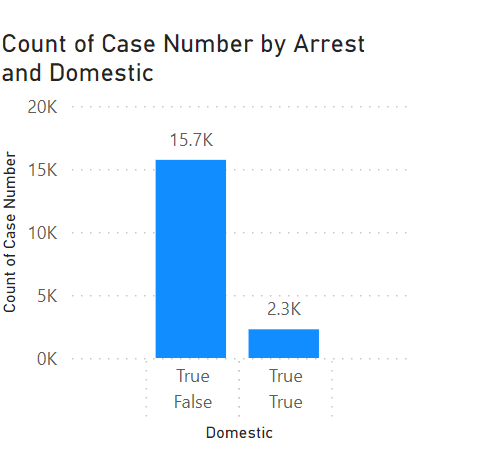
1. Create a monthwise tabular data consisting of two columns, month and total no. of crimes in that month. Also, add one more column where each row of the column contains the total no. of crimes for the previous month. Do we need to use any filter-based DAX function here (All, All except, etc)?



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?

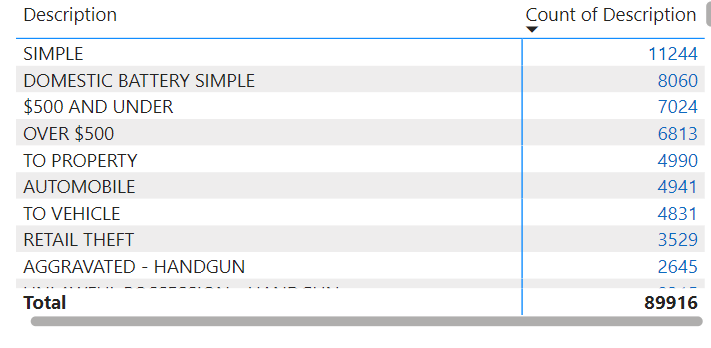
Yes, the arrest rate for domestic crimes is notably low, with only 2,227 out of 15,725 domestic crimes resulting in arrests.

After applying a filter to include only domestic cases by selecting "Yes" in the domestic slicer, I utilized a column chart to plot the number of arrests and domestic cases. The x-axis represents the categories "Arrest" and "Domestic," while the y-axis indicates the count of case numbers.



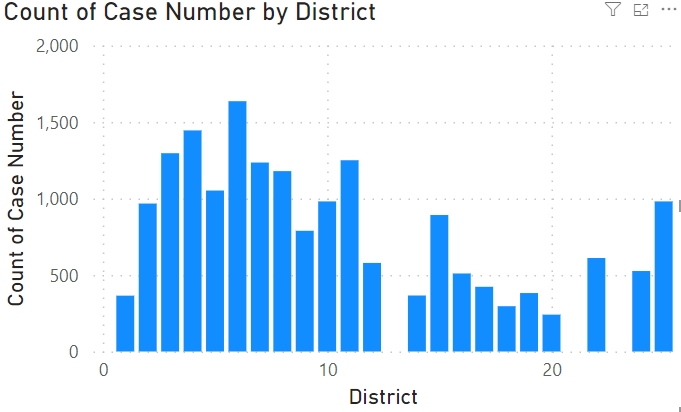
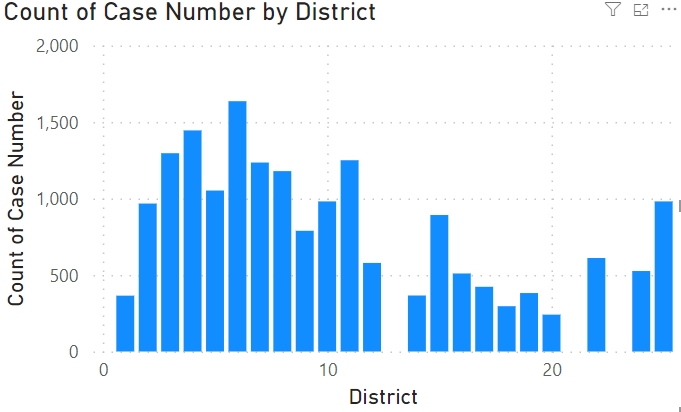
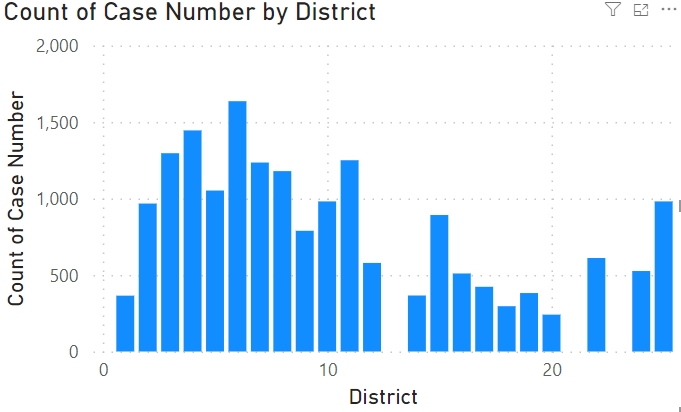
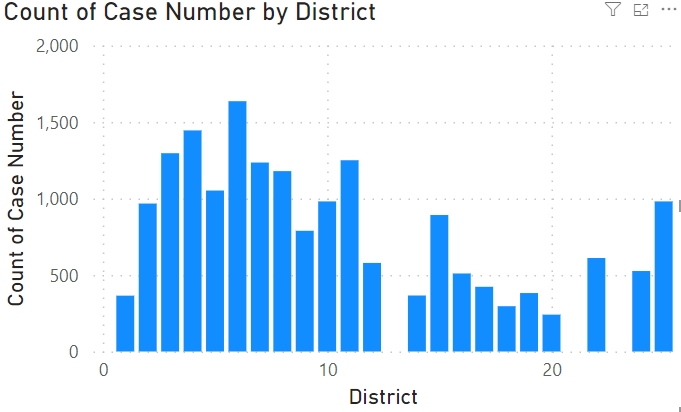
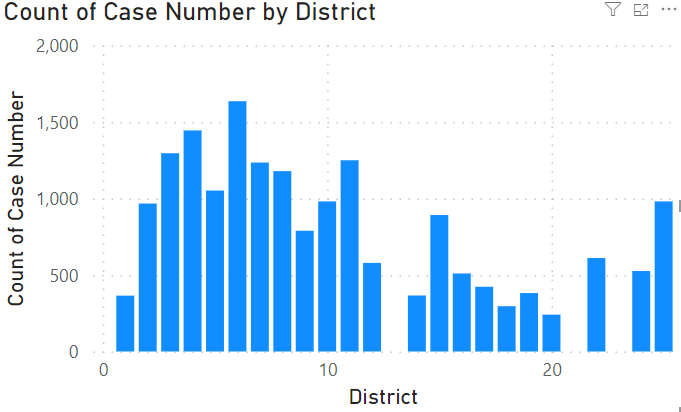
1. Could you generate a visual representation that emphasizes the frequently occurring terms within the "Description" column?

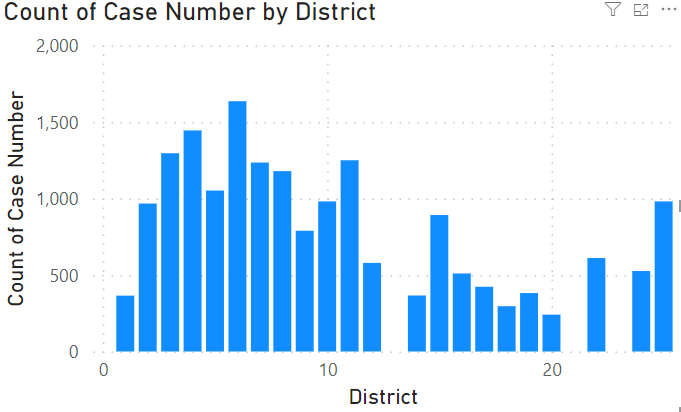
By using matrices, I provided a clear and organized representation of the most frequent descriptions, enabling easy identification of patterns or significant terms within the dataset.



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

After analyzing the bar chart where the district column is plotted on the x-axis and the case number column is plotted on the y-axis, and filtering for domestic crimes only, we observe that districts 3, 4, 6, 7, 8, and 11 have high occurrences of domestic crimes.

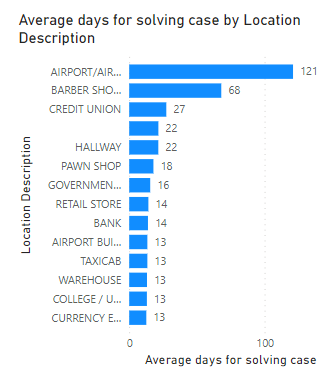
Top of Form



10)Is the solving time of cases also dependent upon the type and locality of crime?

By analyzing the bar chart where the location description is plotted on the y-axis and the average case duration is plotted on the x-axis, we observe that the average days for solving cases vary significantly. For instance, the average duration for resolving cases in an airport/aircraft setting is 121 days, whereas in a pool room setting, it is only 7 days.

Top of Form

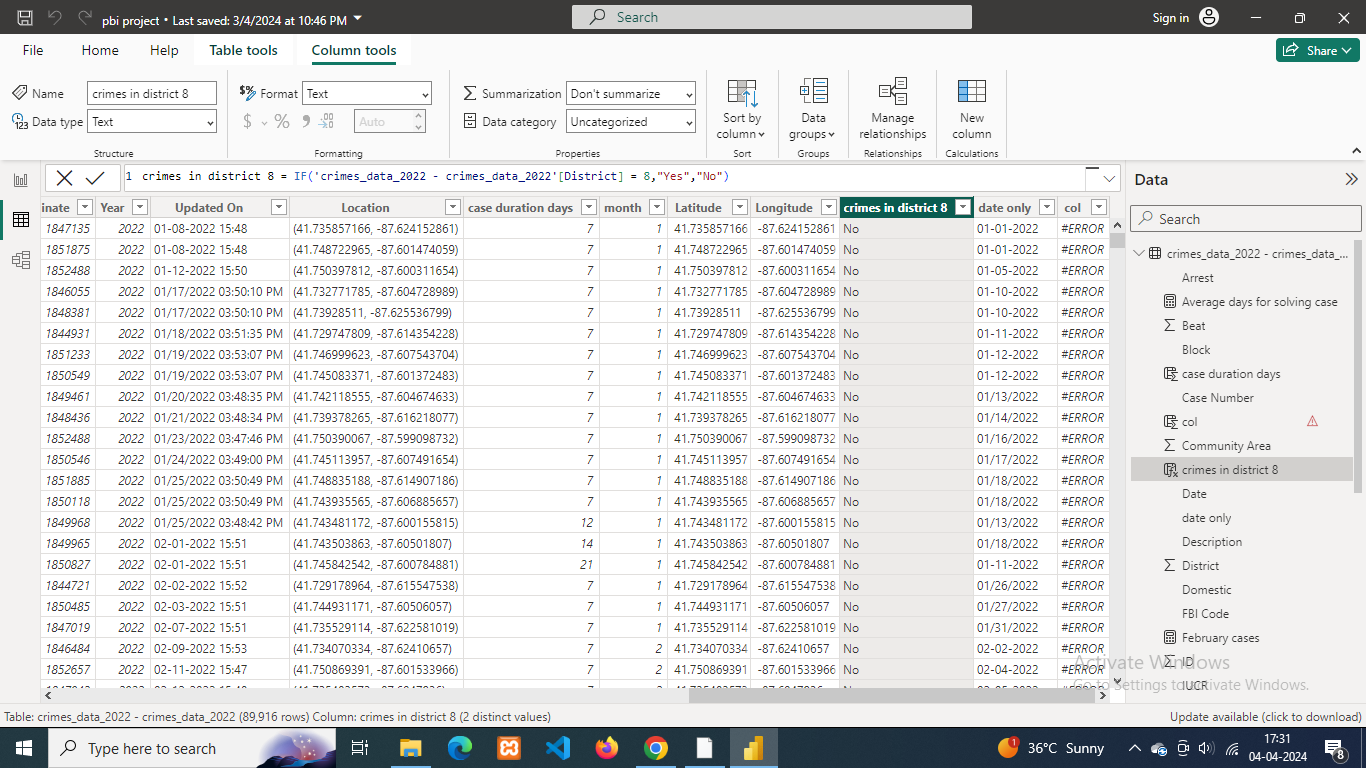


11) Create a calculated column to flag the domestic crimes that took place in District 8.

I used an IF statement to create a calculated column in Power BI. This calculated column flags cases that are registered in District 8. For each row in the dataset, the IF statement checks if the district of the case is District 8. If it is, the calculated column returns a flag indicating that the case is registered in District 8. If not, it returns a no flag.

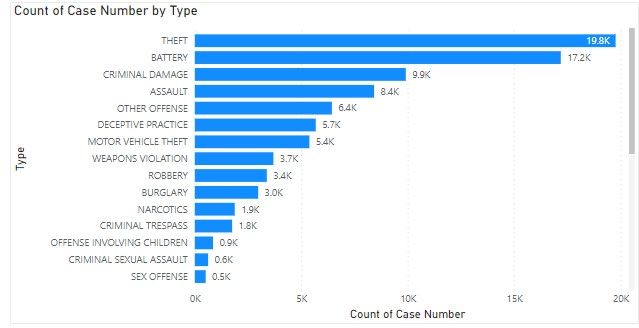
Formula

crimes in district 8 = IF('crimes\_data\_2022 - crimes\_data\_2022'[District] = 8,"Yes","No")



12) 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?

Homicide stands as the most dangerous of all crime categories due to its irreversible nature and the permanent loss of life it entails. Unlike other crimes, homicide results in the ultimate harm—death—impacting not only victims but also families, communities, and society as a whole. Its severity underscores its classification as one of the most serious offenses within the legal system.



13)What do you understand by PowerBI gateway? What are its use cases?

Power BI Gateway is a component of Microsoft Power BI that enables secure data transfer between on-premises data sources and the Power BI service in the cloud. It acts as a bridge between Power BI and your on-premises data sources, allowing you to refresh data in Power BI reports and dashboards that connect to on-premises data.

Here are some key points about Power BI Gateway and its use cases:

* Data Refresh: One of the primary use cases of Power BI Gateway is to enable data refresh for datasets that connect to on-premises data sources. With the gateway installed and configured, Power BI can connect to on-premises databases, files, and other data sources to import or refresh data.
* Scheduled Refresh: Power BI Gateway allows you to schedule data refreshes for your datasets stored in the Power BI service. You can set up refresh schedules to automatically update data at regular intervals, ensuring that your reports and dashboards reflect the most current information.
* Data Security: Power BI Gateway helps maintain data security by establishing a secure connection between on-premises data sources and the Power BI service. It uses encryption and secure communication protocols to protect data during transfer.
* Live Query: Power BI Gateway also supports live query connections to on-premises data sources for real-time data analysis. With live query, Power BI reports can directly query on-premises databases without importing data into the Power BI service.
* Multiple Types of Gateways: There are different types of gateways available depending on your organization's requirements:

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

To address this problem without the provided questions, I would follow a structured approach to data analysis using Power BI. Here's how I would approach it:

* Data Understanding:

Import the dataset into Power BI and understand its structure and contents.

Identify the types of data (categorical, numerical, datetime, etc.).

Explore the relationships between different variables.

* Data Cleaning and Preprocessing:

Address any inconsistencies and missing values in the dataset.

Perform data cleaning operations such as removing duplicates, correcting errors, and handling missing values appropriately.

* Exploratory Data Analysis (EDA):

Conduct exploratory data analysis to understand the distribution and characteristics of the data.

Visualize the data using charts and graphs to identify patterns, trends, and outliers.

* Crime Type Analysis:

Analyze the frequency of each crime type to identify the most prevalent crimes.

Visualize the frequency of each crime type using appropriate charts.

* Arrest Rate Evaluation:

Calculate the percentage of reported incidents that have resulted in an arrest to gauge law enforcement effectiveness.

Visualize the arrest rate over time or by crime type.

* District Crime Distribution Assessment:

Calculate the number of crimes in each district to understand how crime is distributed across the city.

Identify high-crime areas and visualize the crime distribution on a map.

* Domestic Crime Proportion Analysis:

Analyze the ratio of domestic-related crimes to other types of crimes.

Visualize the proportion of domestic crimes compared to other types of crimes.

* Location Description Analysis:

Identify any location descriptions where the number of crimes is higher than expected.

Create a table or visualization to display the frequency of crimes at each location description type.

* Average Time between Reporting and Solving Cases:

Calculate the average time between reporting and solving a case.

Visualize the solving time distribution and analyze any patterns or trends.

* Patrol Area Analysis:

Identify patrol areas where reported crimes were under control.

Visualize crime rates by patrol area and assess effectiveness.

* Calculated Columns:

Create calculated columns for necessary calculations or flags based on the analysis requirements.

Understand the difference between calculated columns and adding columns using Power Query.

* Power Query and M Query:

Utilize Power Query to transform and clean data.

Understand M language syntax and its application in data transformation.

This structured approach ensures comprehensive analysis of the dataset and provides insights into various aspects of crime analysis. If specific questions weren't given, this approach allows for flexibility in addressing the problem and deriving meaningful insights from the data.

15)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?

One district can belong to only one state: This implies that each district in your data set can be associated with only one state. Therefore, it's a one-to-many relationship because each district in your data set (the "many" side) can be related to only one district\_id in the new table (the "one" side), which represents a unique district in a state.

Each district\_id in the new table can have multiple occurrences: Since district\_id is likely to be a unique identifier for districts within a state, it's possible that there are multiple occurrences of the same district\_id in your data set (the "many" side), each corresponding to a different district in your data set. This reinforces the one-to-many relationship.

Therefore, establishing a one-to-many relationship between the district in your data set and the district\_id in the new table would be appropriate for associating districts with their respective state and district identifiers.