**Objective Questions**:

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

Used Power Query Editor to address the inconsistencies & missing values in the dataset.

**Identified and Handled Missing Values:**

* Selected the whole data and replaced null values with “N/A” by using find the replace function.
* Also removed rows from the dataset so as to address the issue of missing values and errors in the data.
* The Ward, X Coordinate, Y Coordinate and Location columns had empty values which were replaced with N/A.

**Addressed Inconsistent Data Types**:

* Use transformation steps in Power Query Editor to clean and standardize data.
* Removed the data types and manually applied the data type of the columns like Date & Updated on columns to Date/Time format to make it feasible for analysis.
* Incorrect data types can lead to unexpected results.

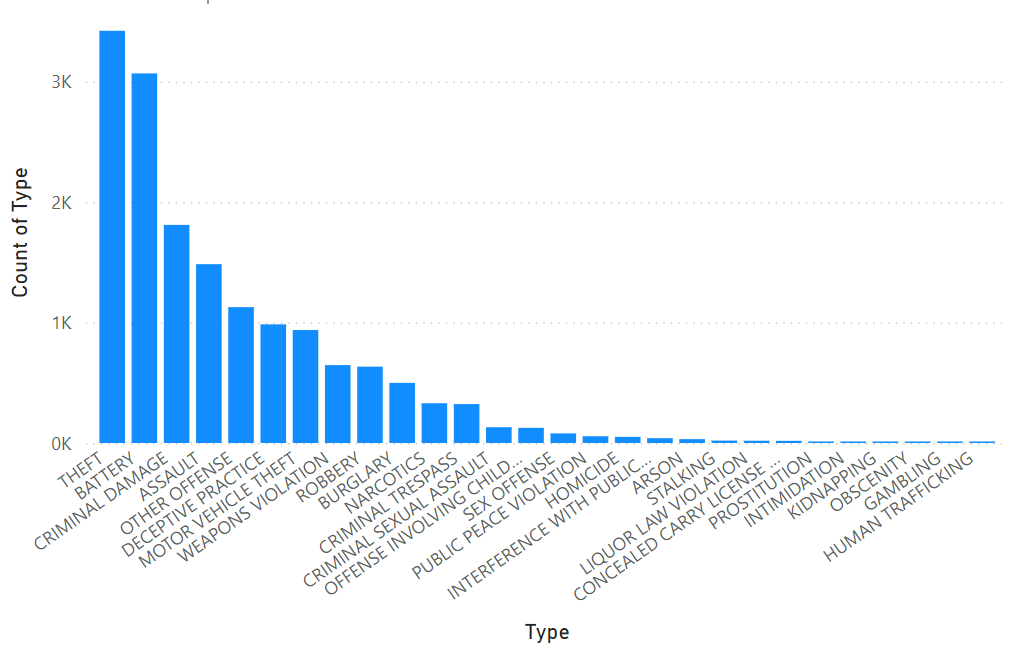
**Removed Duplicates**:

* Duplicate records can skew your analysis.
* Selected all the data and removed duplicates to maintain correctness of the data.
* The number of rows before removing duplicates were 90865 and after removing duplicate rows the count of rows remained 15887. Therefore, 74978 records were duplicates.

-----------------------------------------------------------------------------------------------------------------

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

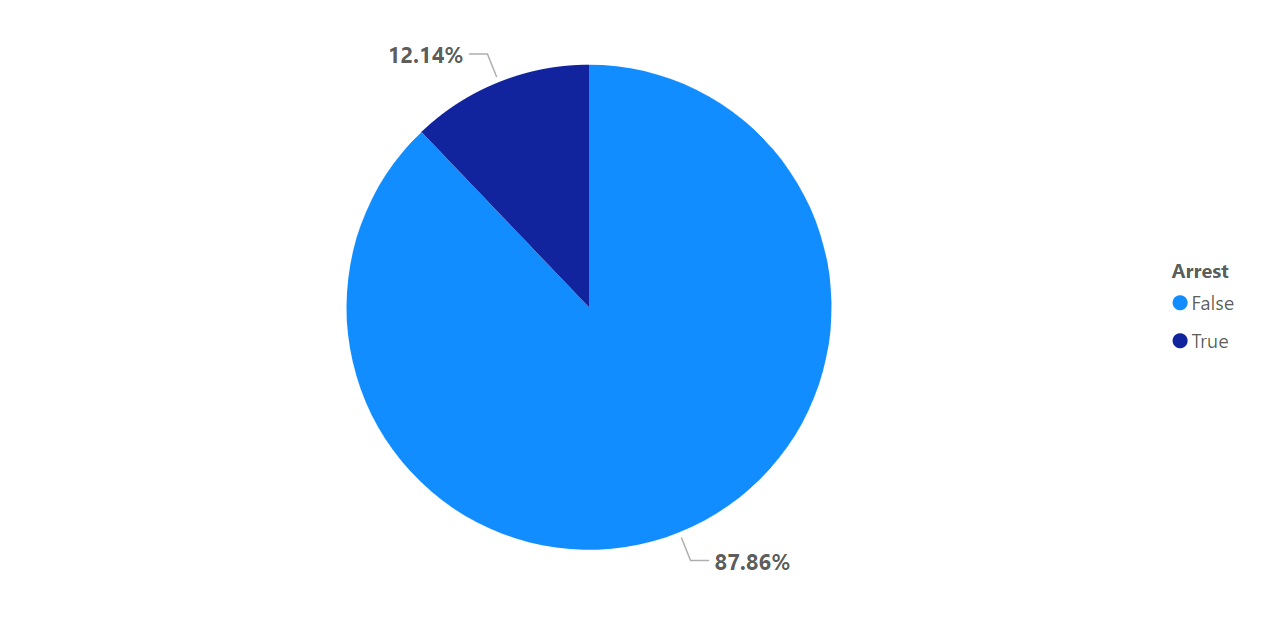
The most prevalent crime type in the area is THEFT based on the frequency of crime type.



-----------------------------------------------------------------------------------------------------------------

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

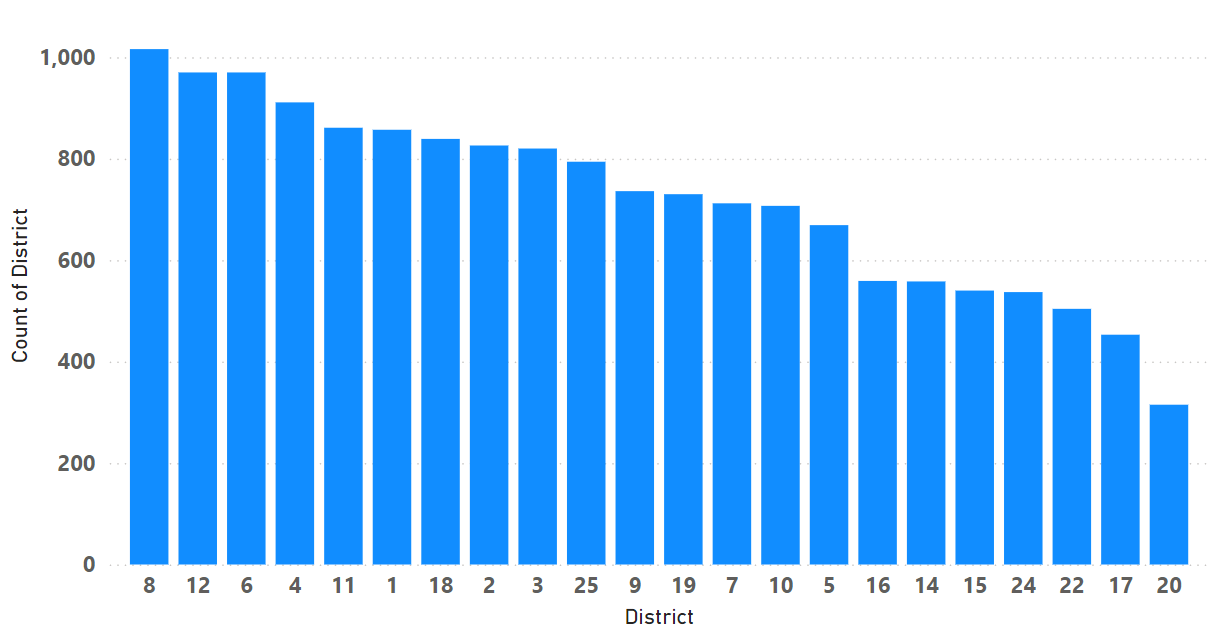
12.14% of reported incidents have resulted into an arrest.



-----------------------------------------------------------------------------------------------------------------

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.

The highest crime area is district number 8.



-----------------------------------------------------------------------------------------------------------------

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

There are 10 categorical columns in the dataset.

-----------------------------------------------------------------------------------------------------------------

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?

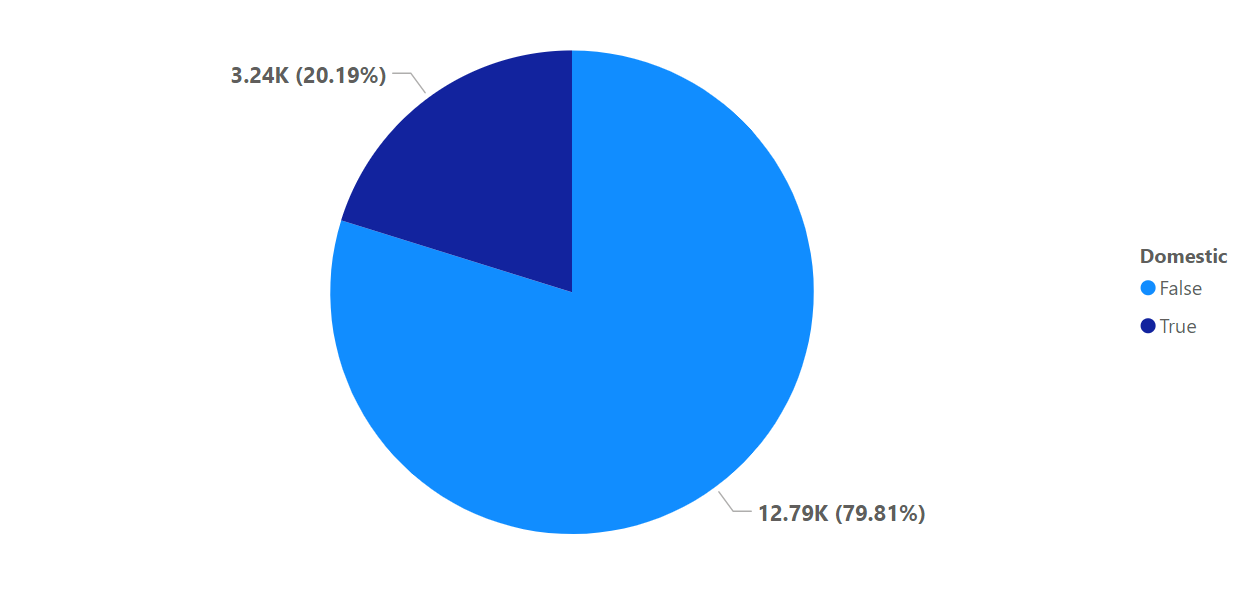
Yes, there were null values in the data. I replaced all the null values from the data with “N/A”. However, the ideal way to handle null values is to

* fill them with mean or median in case of numerical data and mode in case of categorical data.
* Consider null itself as a category.
* Predict it using other non-null columns.

----------------------------------------------------------------------------------------------------------------

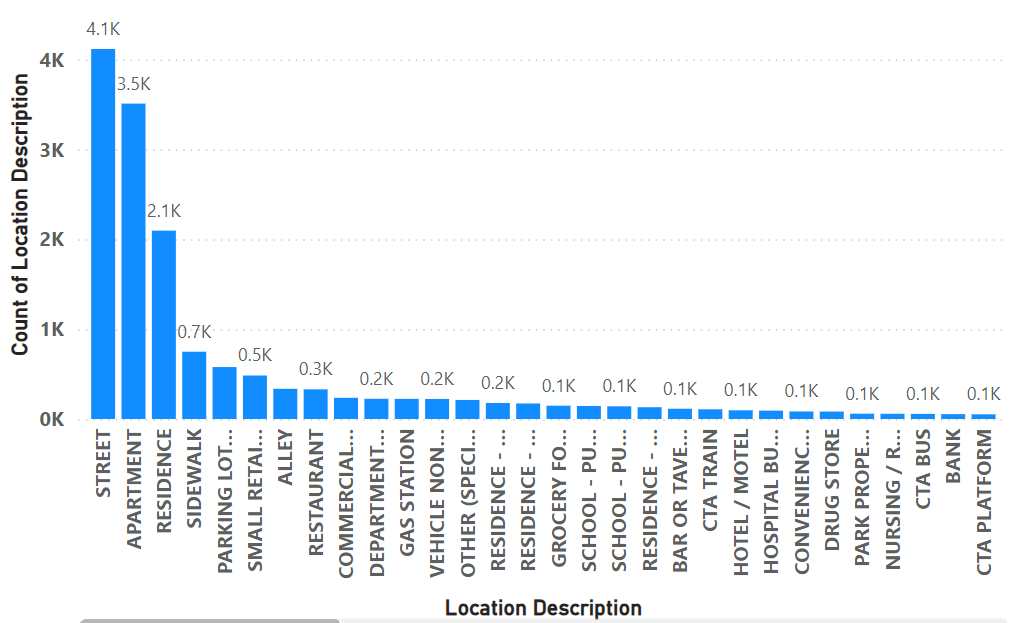
1. Domestic Crime Proportion Analysis: Analyze the ratio of domestic-related crimes to other types of crimes to understand the prevalence of domestic incidents.

20.19% of all the crimes are domestic crimes with a count of 3.24K incidents being recorded by the crime branch.



-----------------------------------------------------------------------------------------------------------------

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.

STREET incidents are mostly in the location description of an incident followed by incidents in APARTMENTS. 

-----------------------------------------------------------------------------------------------------------------

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

It takes around 201.64(~ 202) days to solve a case.

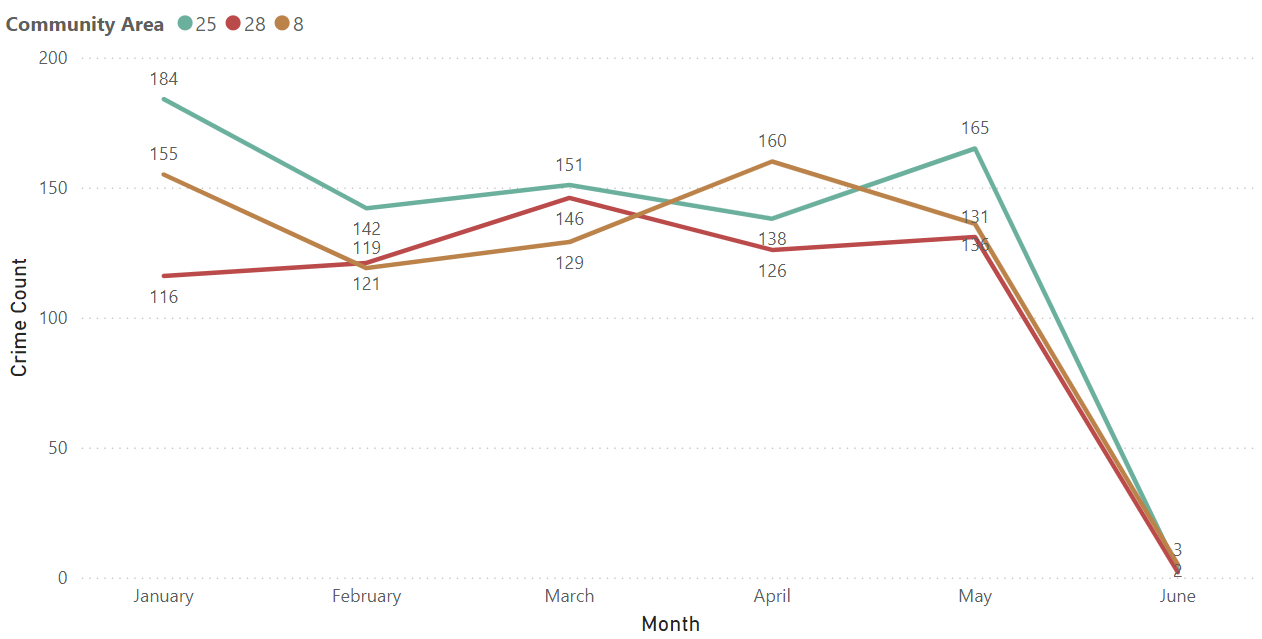
Formula used: - Average Days in Solving Case = AVERAGEX('crimes\_data\_2022 - crimes\_data\_2022', DATEDIFF('crimes\_data\_2022 - crimes\_data\_2022'[Date],'crimes\_data\_2022 - crimes\_data\_2022'[Updated On],DAY ))

---------------------------------------------------------------------------------------------------------------------------

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

**Criteria**:

* For the above objective question, I analysed the change in number of crimes in different community area over months.
* Used a Line chart to visualize the change in crime count.
* Used a filter to get only the Top 3 Community areas where the number of crimes reported were the highest.
* Also used a filter to get the data from January – June only because maximum variation in crime count is shown in this time period only.



It can be clearly seen from the chart that the petrol officers of area 25, 28 & 8 should be rewarded because the decrease in crime rate is maximum in these 3 community areas.

-----------------------------------------------------------------------------------------------------------------

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

No, I have not created a calculated column in the project till now.

**Calculated Column**: A calculated column is a new field that you create within a table in your Power BI model. It’s based on a Data Analysis Expressions (DAX) formula that you define. Unlike measures (which are used in visualizations), calculated columns are part of the underlying data model.

* + **Purpose**: Calculated columns allow you to create new data points by performing calculations row by row. These columns are computed during data loading and are stored in the model.
  + **Where to Create**: You can add a calculated column in Data View, Report View, or Model View.
  + **Use Cases**:
* Deriving new attributes: For example, calculating profit margin by subtracting costs from revenue.
* Creating flags or categories based on conditions: E.g., categorizing customers as “High Value” or “Low Value.”
* **Performance Considerations**:
* Calculated columns consume RAM and increase the file size.
* They don’t require a full refresh of the table when added.
* Useful for scenarios where the calculation is row-context-dependent.

**Custom Column (Computed Column in Power Query)**: A custom column (also known as a computed column in Power Query) is created during data transformation using Power Query Editor. It’s based on M language (the language behind Power Query).

* **Purpose**: Custom columns allow you to transform data before it’s loaded into the model. They’re part of the query folding process.
* **Where to Create**: You add custom columns exclusively in Power Query Editor.
* **Use Cases**:
* Cleaning and transforming data: E.g., extracting parts of a text column, converting data types, or merging columns.
* Creating lookup tables or additional attributes.
* **Performance Considerations**:
* Custom columns require a full refresh of the table when added.
* They don’t consume RAM but affect query execution time.
* Useful for scenarios where the calculation is context-independent.

-----------------------------------------------------------------------------------------------------------------

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

There are 193 criminal cases reported for type “THEFT” in the district 8.

Formula used: - Crimes\_theft\_District8 = CALCULATE(COUNTX('crimes\_data\_2022 - crimes\_data\_2022','crimes\_data\_2022 - crimes\_data\_2022'[ID]),'crimes\_data\_2022 - crimes\_data\_2022'[Type] = "THEFT",'crimes\_data\_2022 - crimes\_data\_2022'[District] = "8")

-----------------------------------------------------------------------------------------------------------------

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

Yes, we can easily separate the Longitude and Latitude from the Locations column by transforming data using Power Query Editor and then splitting the column by using the delimiter option. Then to split use Comma(,) as the delimiter.

-----------------------------------------------------------------------------------------------------------------

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?

The code that appears in the formula bar is

Table.AddColumn(

table as table,

newColumnName as text,

columnGenerator as function,

optional columnType as nullable type

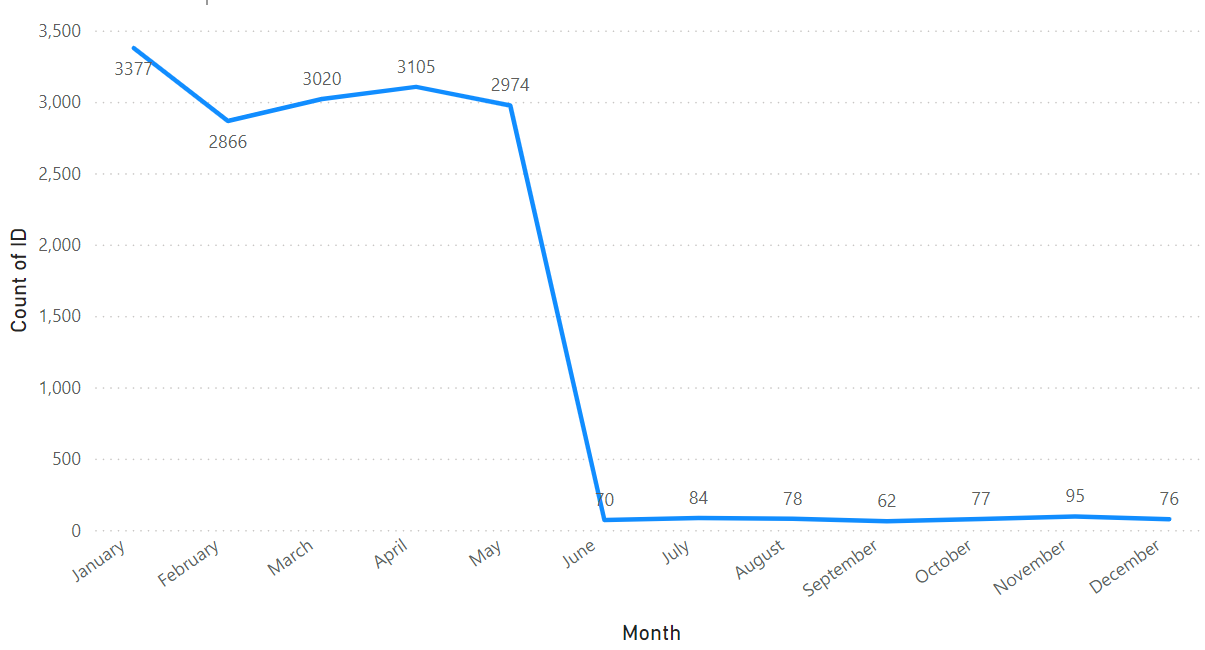
) as table

M Query: - M is a functional language used for data transformation and data preparation within Power Query. M allows you to filter, combine, and transform data from different sources, creating a unified dataset.

**Subjective Question:**

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

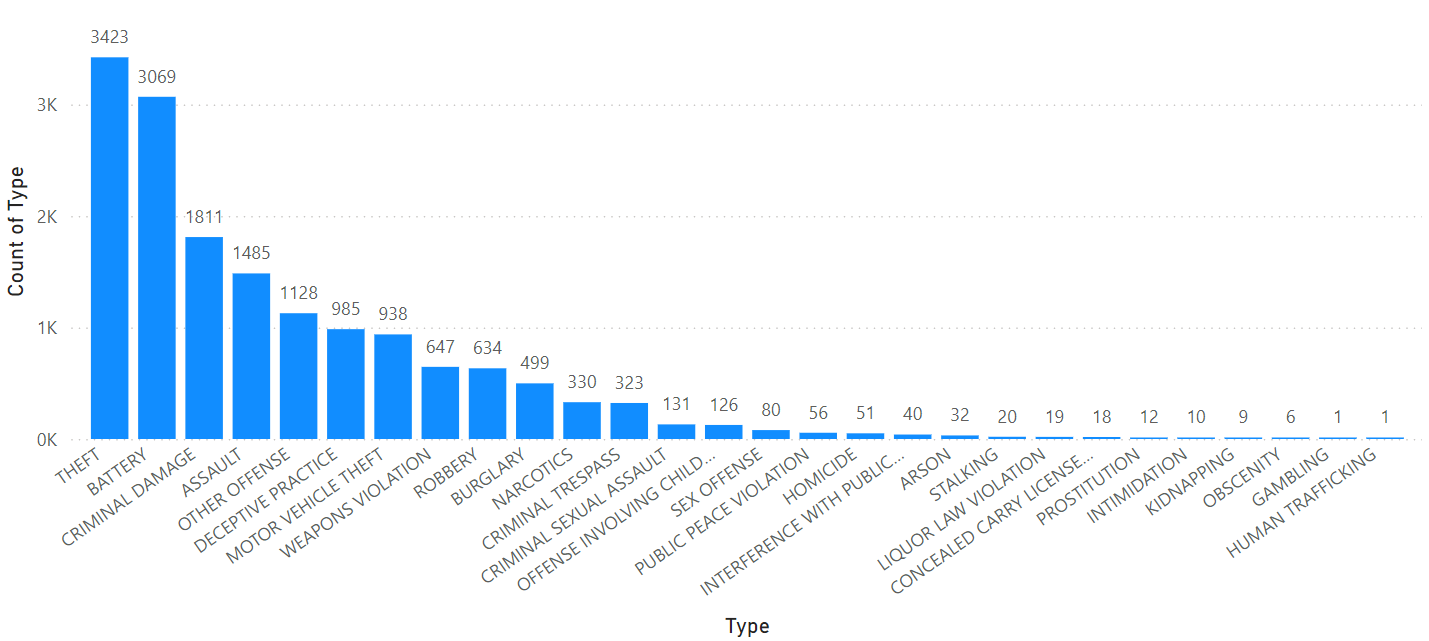
There is a drastic decrease in the number of crimes between the month of May & June. The operation has generated results and there is decrease in the number of crimes in the second half of the year.



---------------------------------------------------------------------------------------------

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 type of crime we should focus is none other than THEFT because the incidents of theft recorded are very large in number. Also, we should decrease the incidents of the type BATTERY to have a huge improvement in the number of crimes.



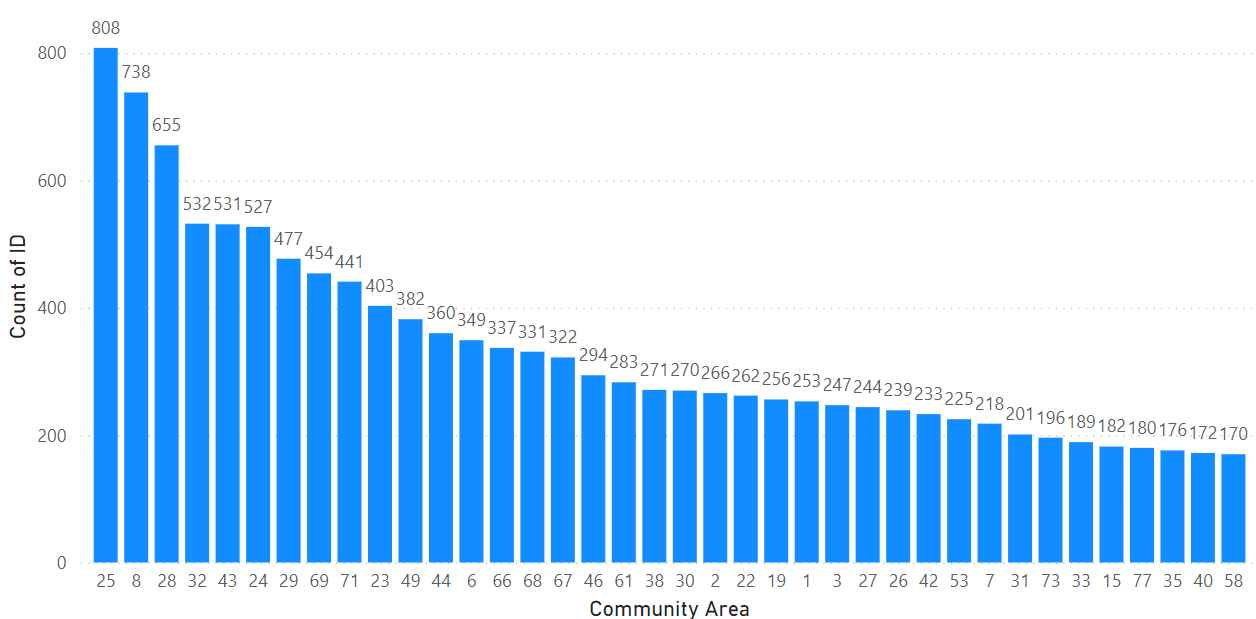
Recommendations: - The methods to have an improvement in the number of crimes are –

* Deploy a greater number of patrol officers near residential & commercial areas to decrease the incidents of theft.
* Patrolling should be done at regular intervals.
* Stricter actions should be taken against the criminals.

------------------------------------------------------------------------------------------------------

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

The community area number 25, 8 & 28 have recorded higher number of cases filed with maximum 808 cases in community area 25 according to the data.



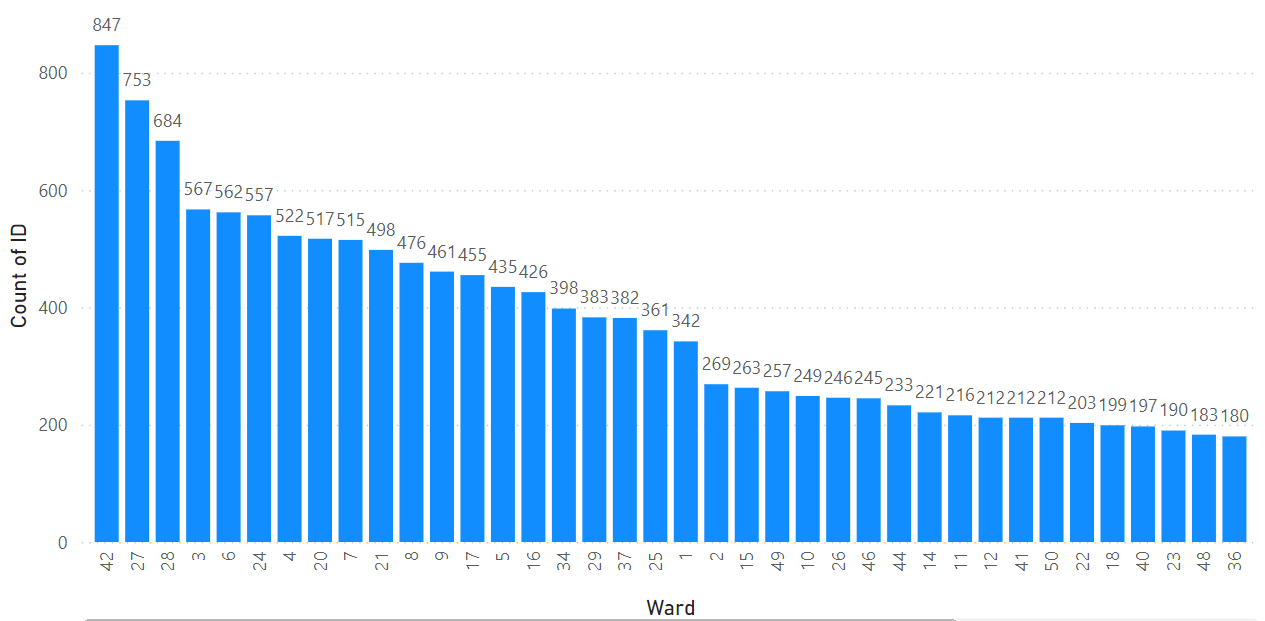
Recommendations: - Measures to reduce these number of crimes are –

* CCTV should be installed at regular intervals in the area.
* Vigilant patrolling officers should be appointed frequently.
* Strict punishments to the criminals to set an example.

----------------------------------------------------------------------------------------------------------------

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

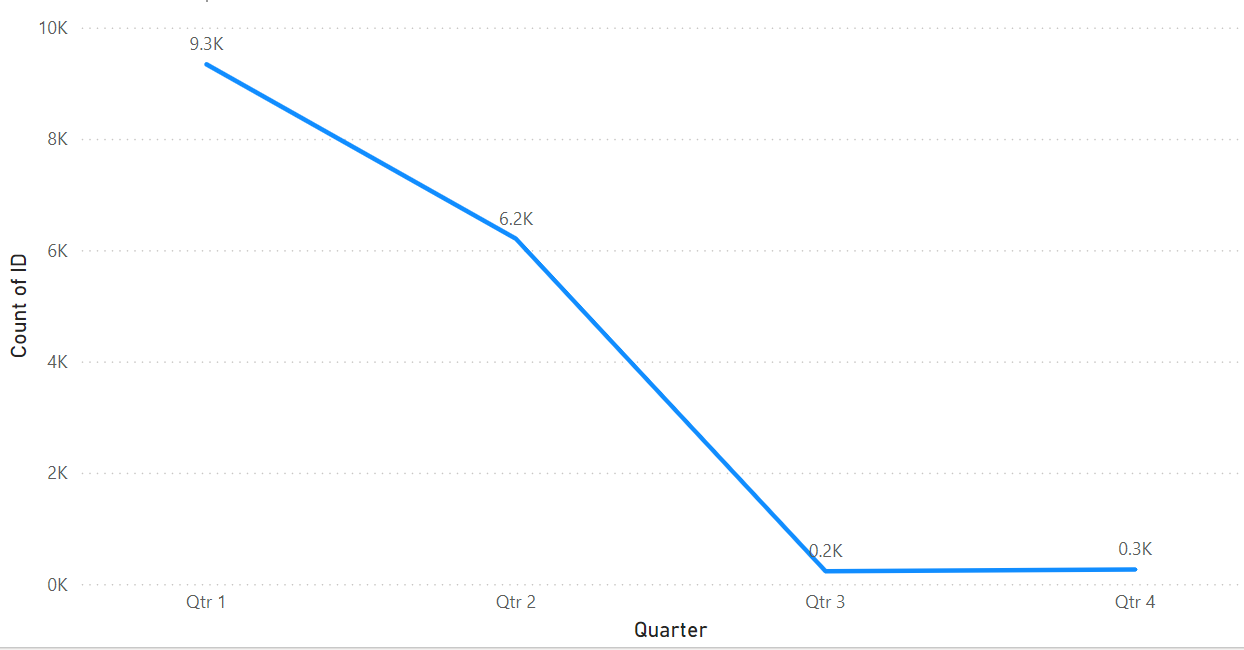
Ward number 42, 27 & 28 have higher number of cases registered therefore security improvements should be made to reduce crime.



----------------------------------------------------------------------------------------------------------------

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

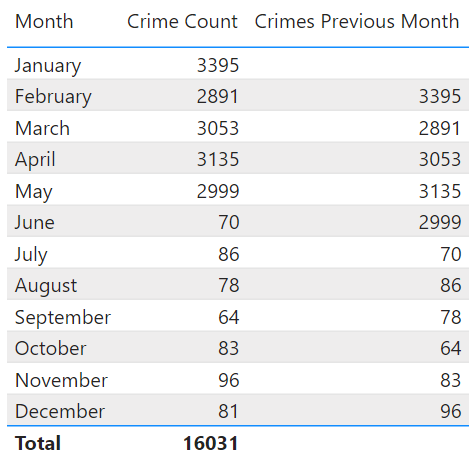
The Crime Rate over Time analysis shows that there is a drastic drop in the number of cases from 1st quarter to 4th quarter. It shows that the hard work of crime department had paid off.



Insights: - There is one nuance with the dataset that the cases registered have only the first 6 days of each month as the date and not after that. This results into a doubtful condition over the correctness of the data.

---------------------------------------------------------------------------------------------------------------

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)?

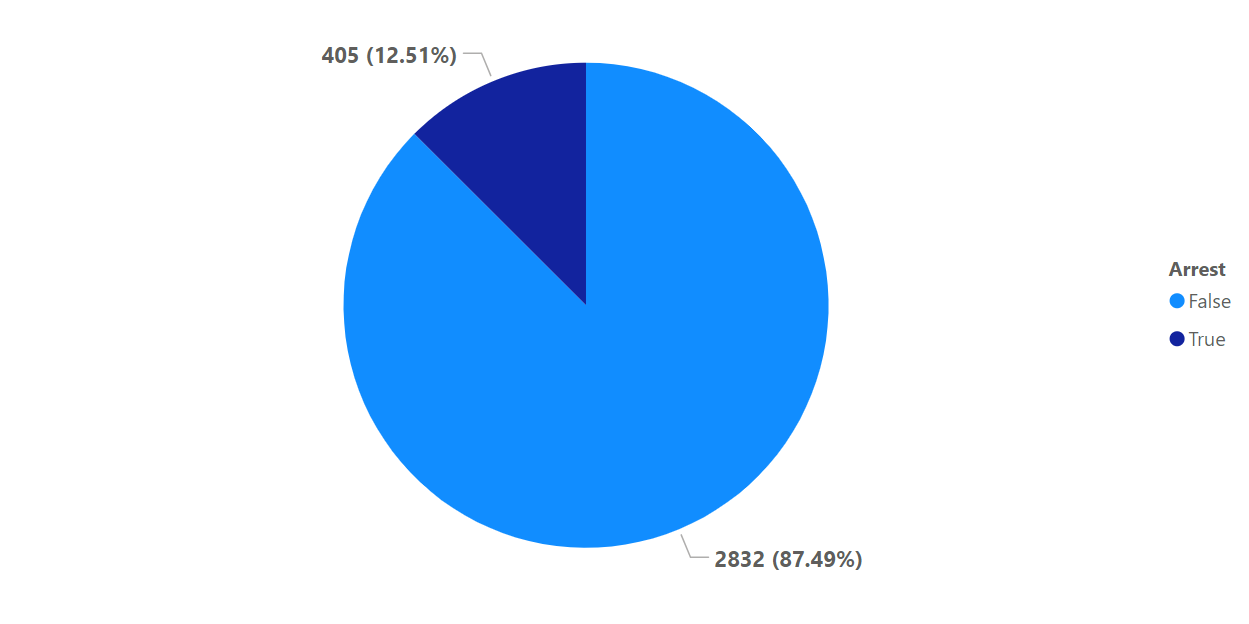


A new measure called ‘Crimes Previous Month’ was created using CALCULATE Function with PREVIOUSMONTH & ALL EXCEPT.

-------------------------------------------------------------------------------------------------------

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?

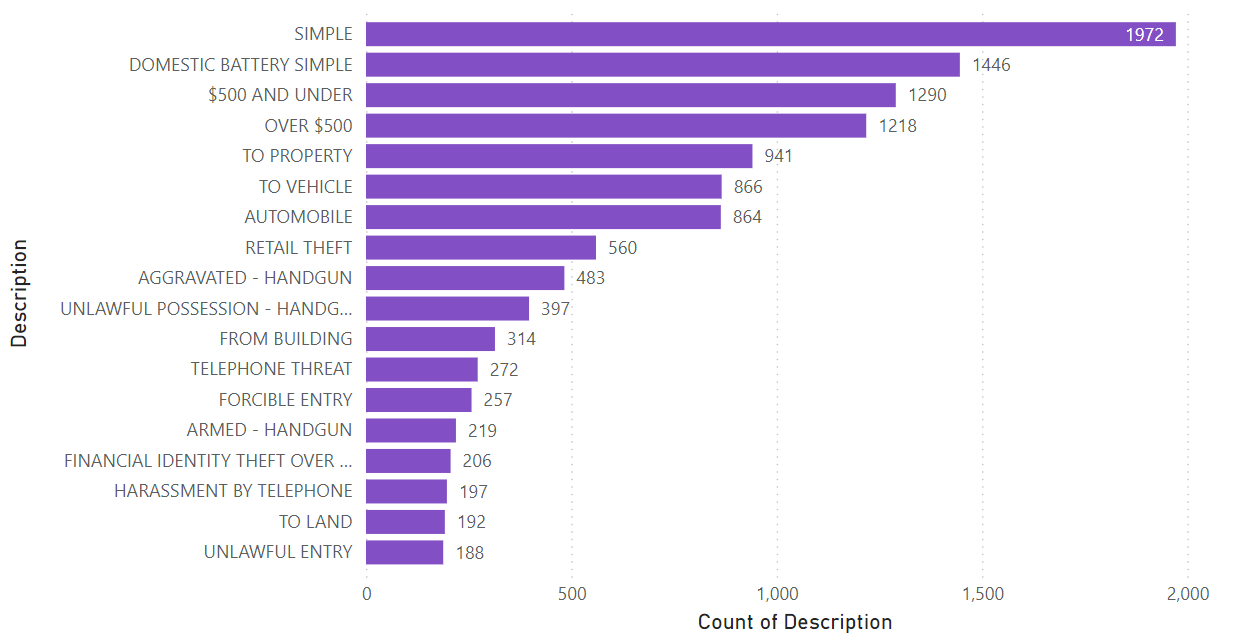
Yes, this trend is clearly visible in our data because only 405 domestic crime cases led to an arrest out of the total domestic crime cases.



---------------------------------------------------------------------------------------------

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

Yes, it can be generated using the Bar Chart.



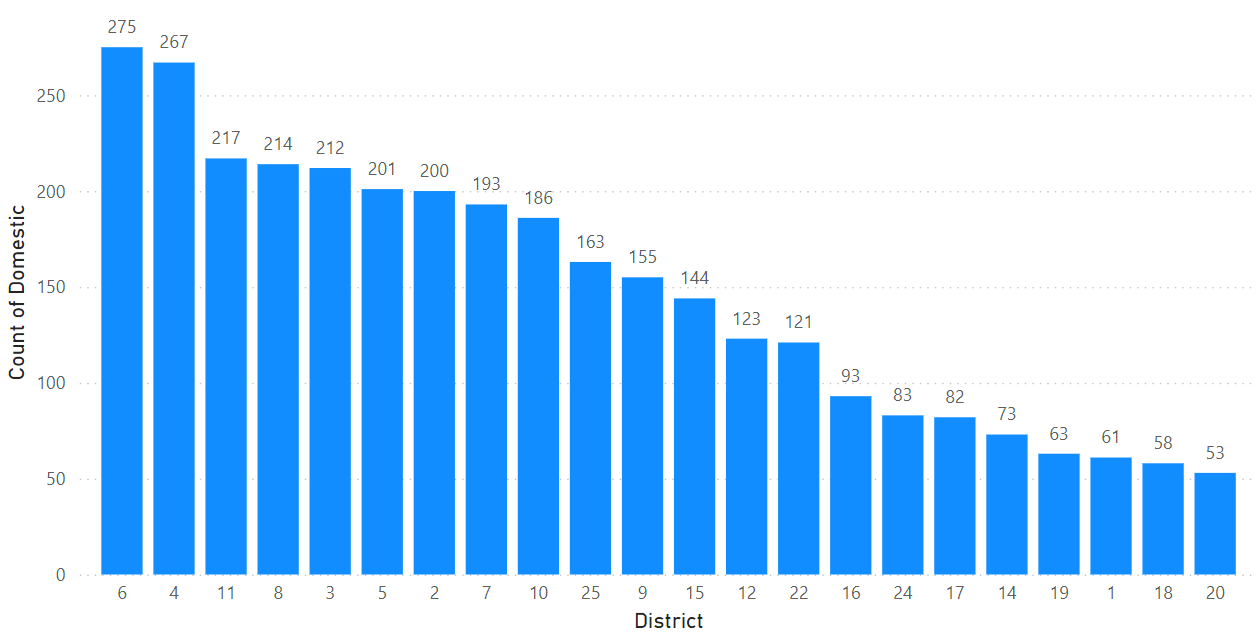
It can be seen that the term ‘SIMPLE’ has occurred the greatest number of times in the crimes reported followed by ‘DOMESTIC BATTERY SIMPLE’ & ‘$500 AND ABOVE’.

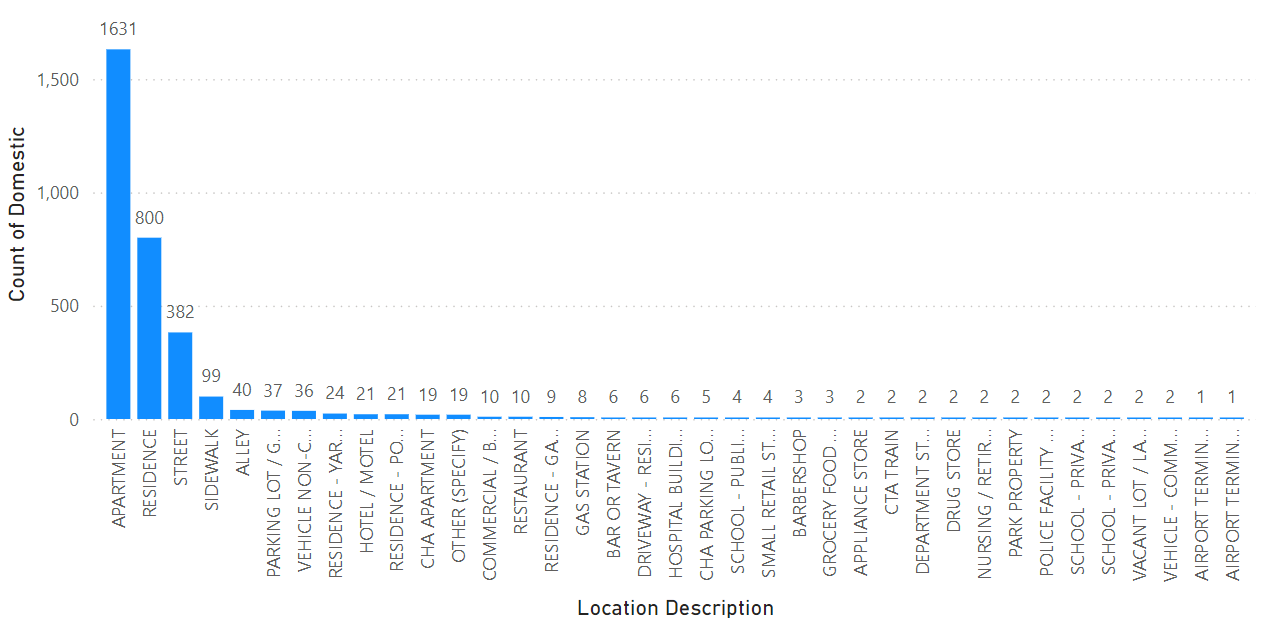
-------------------------------------------------------------------------------------------------------

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

Yes, there are districts where number of domestic crimes reported are high as compared to other districts like District 6 and 4.

Also, the number of domestic crimes reported in apartments in way more than in residence and streets.





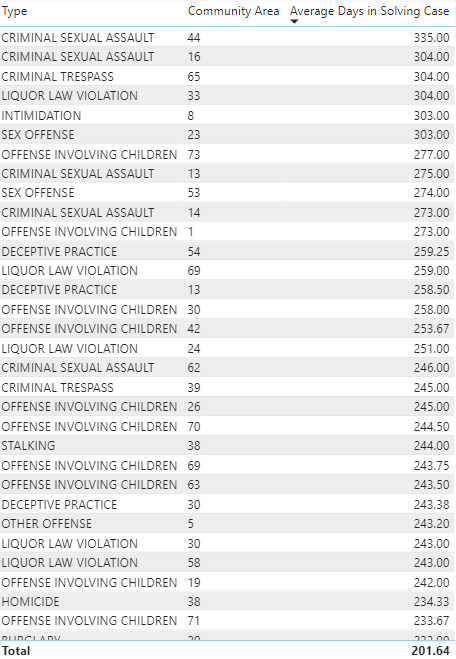
Recommendations: - There can be various measures that can be taken to resolve this problem like –

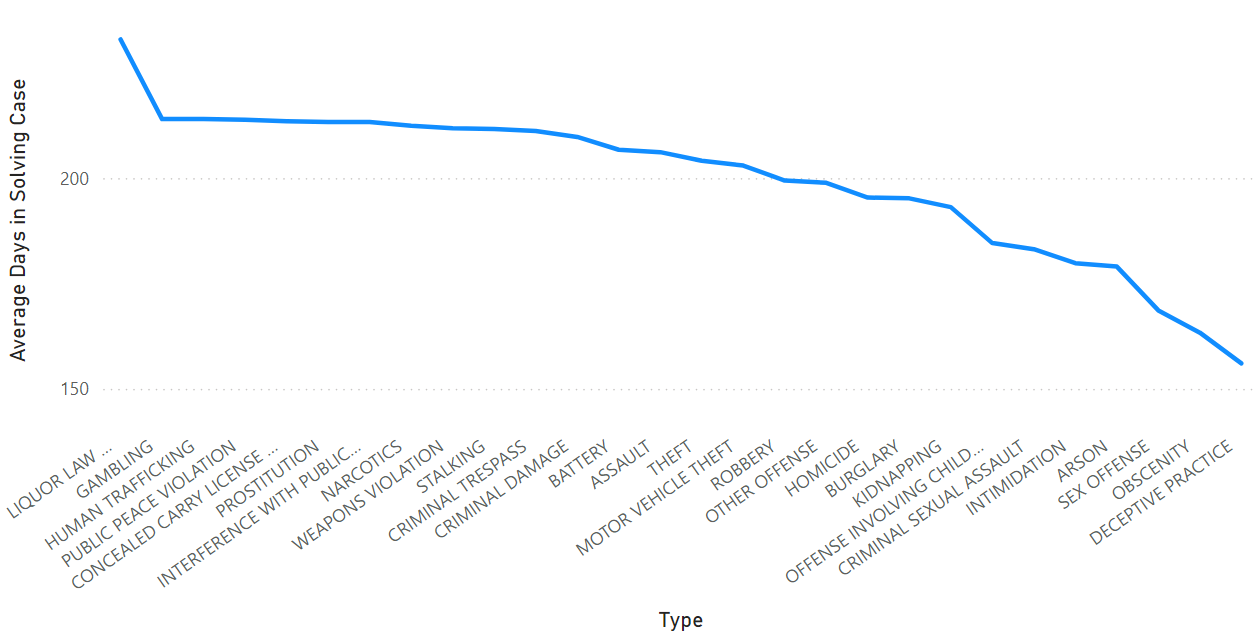
* Starting campaigns to create awareness among people.
* Making people familiar with self-defence techniques.
* Resolving these issues at a faster pace to gain trust of the people.

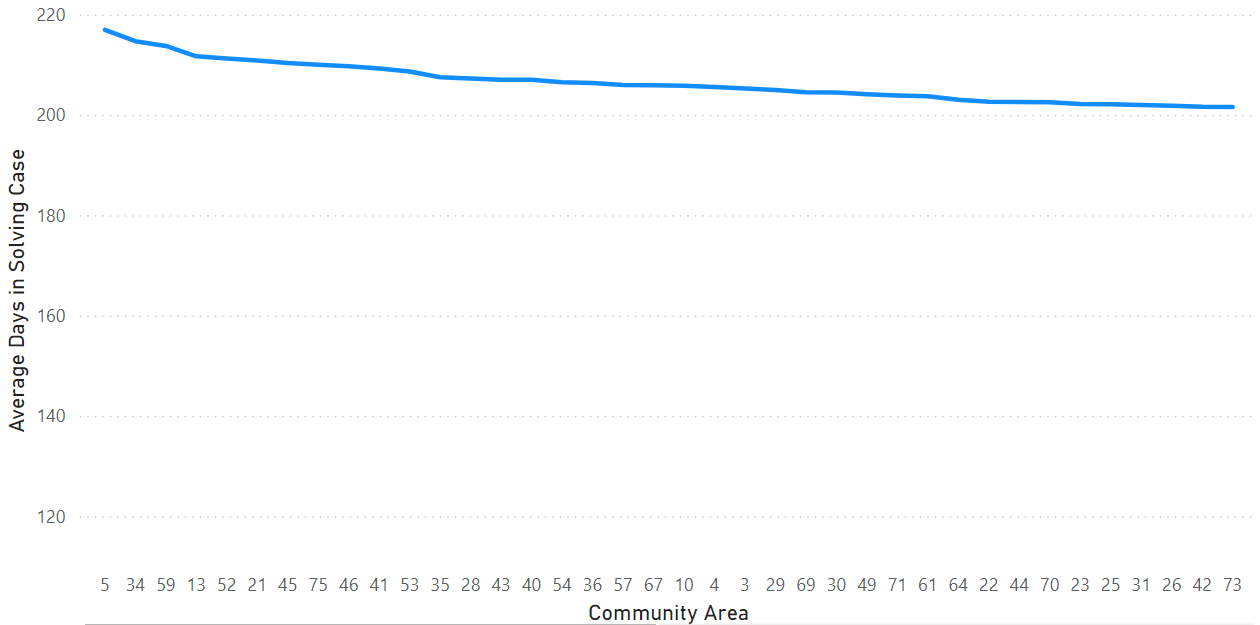
----------------------------------------------------------------------------------------------------------------

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

There is no such pattern which shows that the solving time of cases depend on the type and locality of crime.







Insights: - It can only be seen that the Liquor Law Violation crimes cases take a lot of time to be solved. However, the number of days to solve a case in almost all the community areas is around 190 – 215 days.

----------------------------------------------------------------------------------------------------------------

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

Formula Used: -

Domestic Crimes in District 8 = IF('crimes\_data\_2022 -crimes\_data\_2022'[District] = "8" && 'crimes\_data\_2022 -crimes\_data\_2022'[Domestic] = True, "Flag")

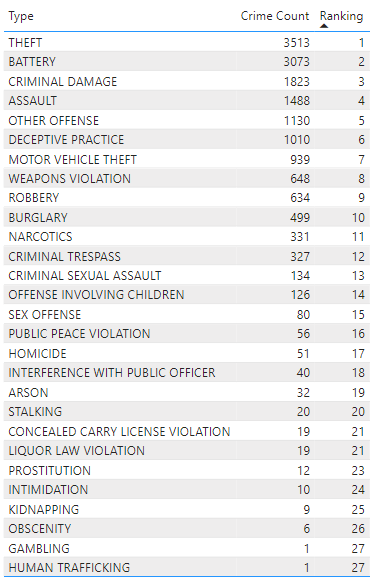
----------------------------------------------------------------------------------------------------------------

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?

According to me Criminal Sexual Assault, Homicide, Human Trafficking, Narcotics & Kidnapping are equally dangerous to the society.

Formula Used: -

Ranking = RANKX(ALL('crimes\_data\_2022 - crimes\_data\_2022'[Type]), 'crimes\_data\_2022 - crimes\_data\_2022'[Crime Count],,DESC)



----------------------------------------------------------------------------------------------------------------

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

Power BI Gateway is a crucial component that bridges the gap between your on-premises data sources and the cloud-based Power BI service. It allows you to securely connect to data stored within your organization’s network, even if that data resides behind firewalls or in private networks.

Purpose and Use Cases: -

* + On-Premises Data Connectivity: The primary purpose of a Power BI Gateway is to enable connectivity to on-premises data sources. If your data lives within your organization’s network (e.g., SQL Server databases, file shares, or other proprietary systems), you’ll need a gateway.
  + Automatic Data Refresh: When you create Power BI reports or dashboards that rely on on-premises data, the gateway ensures that these reports stay up to date. It facilitates scheduled data refreshes, pulling in fresh data from your local sources into Power BI Service.
  + DirectQuery Support: If you’re using DirectQuery mode (where queries are sent directly to the data source), the gateway plays a critical role. It allows Power BI to query your on-premises database in real-time.
  + Security Isolation: Sometimes, you might want to isolate certain connectors or functions for security reasons. The gateway provides this isolation layer, ensuring that only authorized users can access specific data sources.
  + Multiple Cloud Services: Beyond Power BI, the same gateway can be used by other Microsoft cloud services like Power Apps, Power Automate (formerly known as Flow), Azure Logic Apps, and Azure Analysis Services. This consolidation simplifies management.

----------------------------------------------------------------------------------------------------------------

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

If the objective and subjective questions weren’t given, I would have first gone through the dataset provided to understand the key metrics required to perform this analysis. Then I would have framed questions around those metrics to see the patterns and trends around them. The questions would be like: -

* Which type of crime has the highest occurrence in a particular region?
* What is the average time in solving a case in every district?
* Which district has the greatest number of crimes?
* What percentage of criminals were arrested out of the total?

And many more… These questions will help me to gain insights from the data and presented the information in form of visualizations.

----------------------------------------------------------------------------------------------------------------

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?

There would be a many-to-one relationship between the district column in our data and the district\_id column of the new table.

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***