**Learners have to develop a Report to support the answers to the following questions and suggestions**

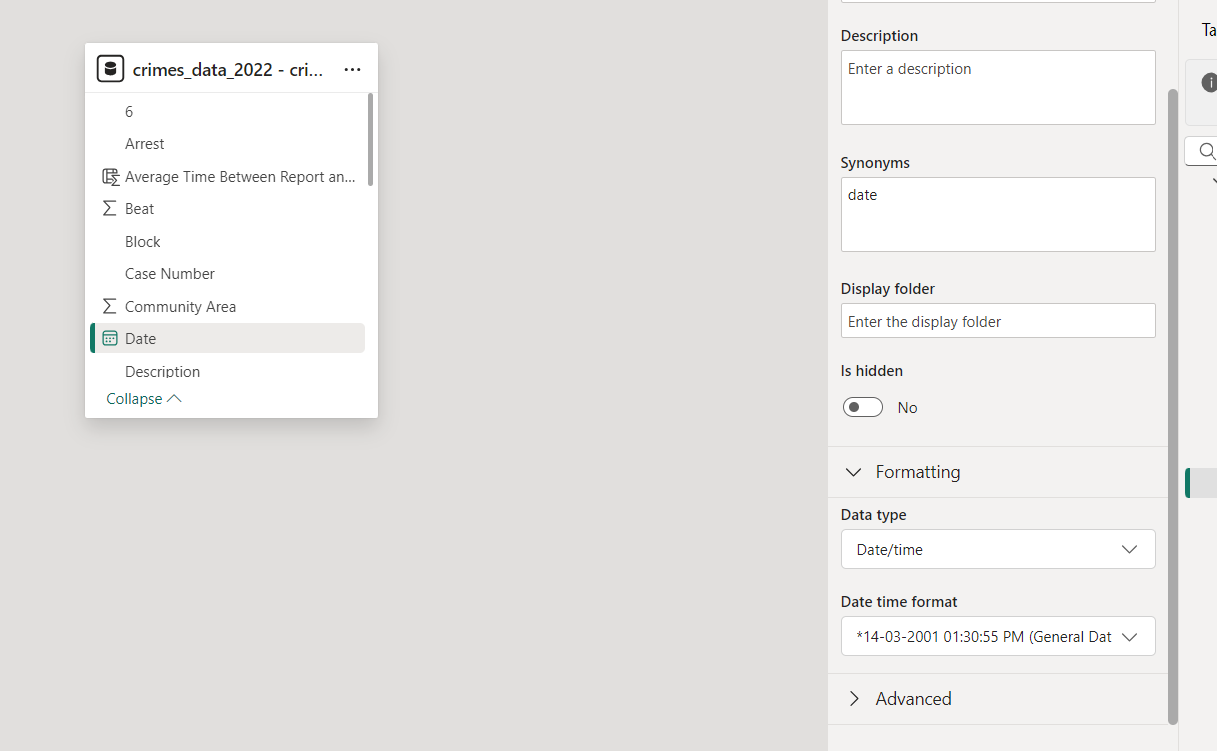
**Objective Questions**:

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

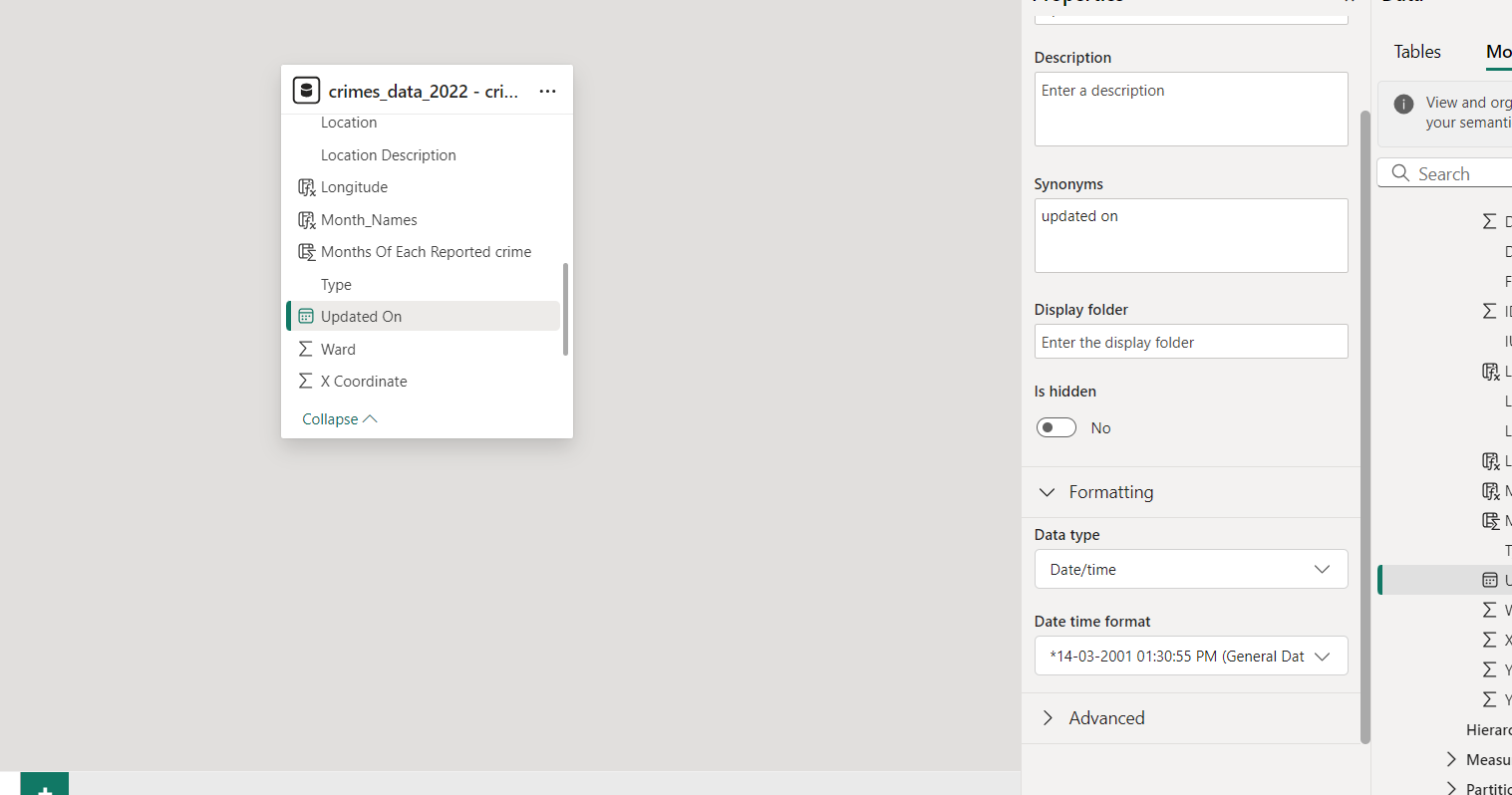
=

1. I took the data in power query editor and removed the blank rows from the option of remove rows.
2. I replaced null in text column with ‘Not Available’ text.
3. I removed the duplicate rows as well from the dataset.
4. I replaced “/” with “-“ of date and updated on column.
5. And changed the data type of date and updated on column from model view as it was throwing error while trying to change it from power query editor.

Date – Datatype Change.



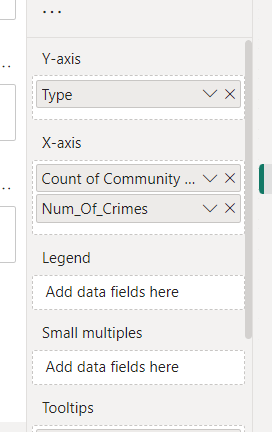
Updated – On – Datatype Change.

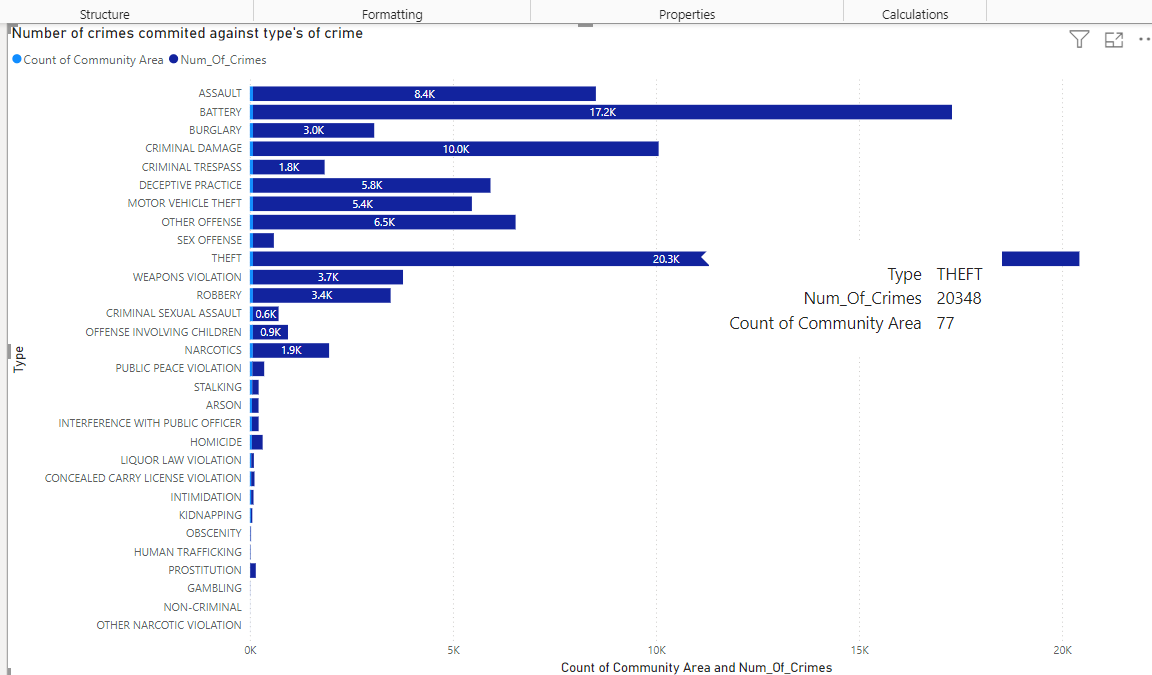


1. Crime Type Analysis: Assess the frequency of each crime type to identify the most prevalent crimes occurring in the area.

=

* The column ID represents each committed crime uniquely.
* And distinct count of community area represents the count of all distinct areas.
* Theft is the most prevalent crimes occurring in the area.

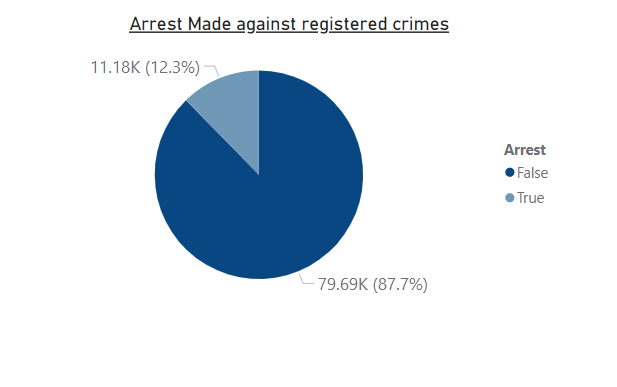




1. Arrest Rate Evaluation: Analyze the percentage of reported incidents that have resulted in an arrest to gauge law enforcement effectiveness.

=

* By creating a pie chart for count of id in regards of arrest (True or False)
* We got the percentage of reported incidents that have resulted in an arrest.
* The percentage of arrest is – 12.3%



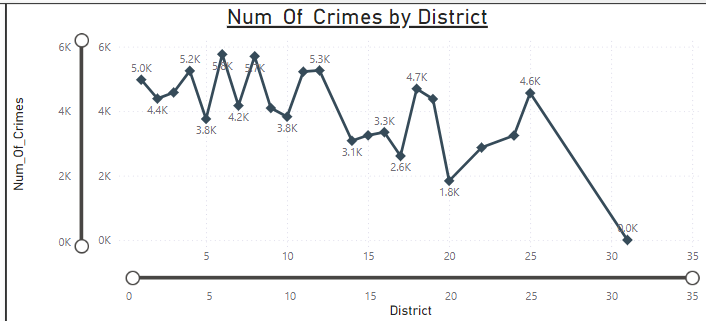
1. District Crime Distribution Assessment: Calculate the number of crimes in each district to understand how crime is distributed across the city and identify high-crime areas.

=

* We created a line chart to calculate the number of crimes in each district for viewing how crime is distributed across the city and identify high-crime areas.
* The districts that have the high-density crime areas –

**District – Crimes**

* 6 – 5743
* 8 – 5670
* 11 – 5200
* 4 – 5214
* These are the district with high density crime areas.



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

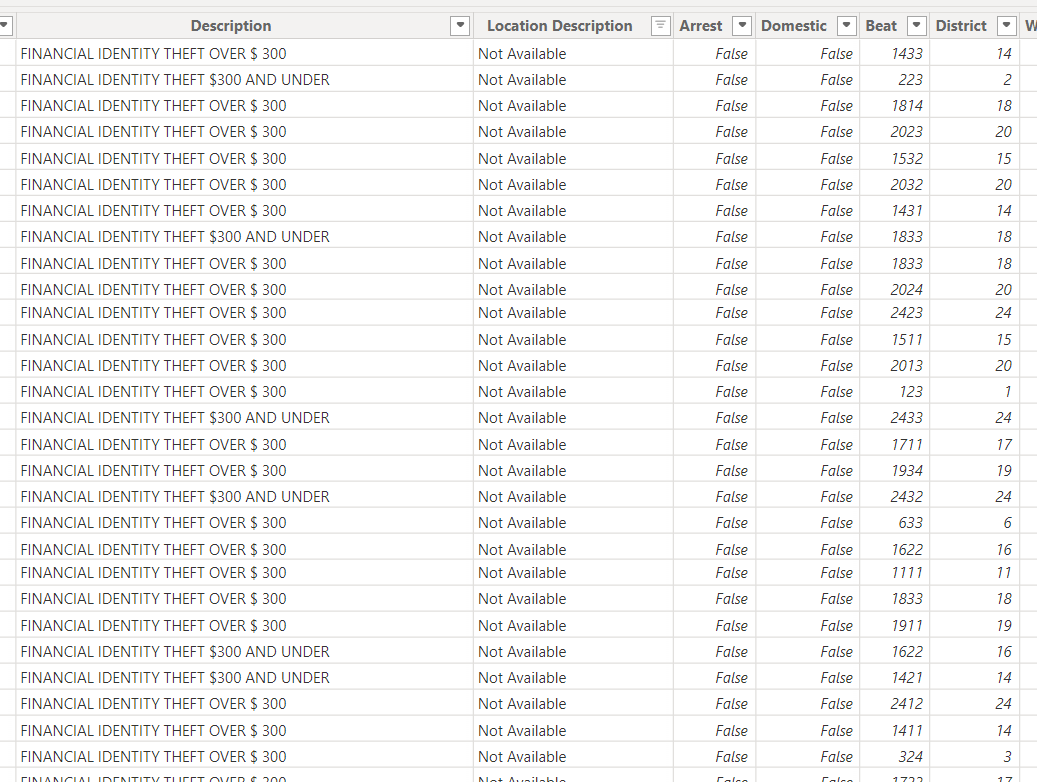
=

* We can categorise it by Type, District, Location Description, Arrest, Domestic, Ward, Community Area, FBI Code.
* In total we have 8 categorical attributes in 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?

=

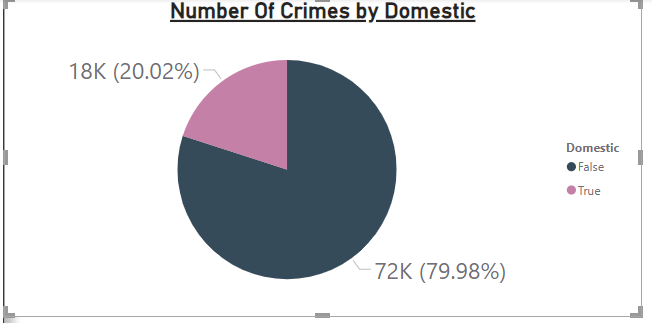
* There was null value in the data.
* I unselected the null values from the filter option in the power query editor.
* The ideal way to handle the null values will be removing those from the source file itself.



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

=

* As the domestic column contains only True and False text.
* We created one pie chart and against domestic we took number of crimes.
* We got 18K (20.02%) of domestic crime and 72K (72.98%) of other crimes.
* So, the ratio of domestic crime to other crime is 1:4.

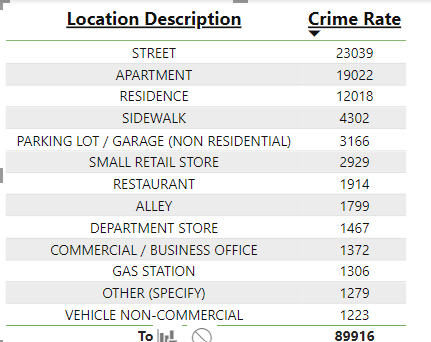


1. Is there any “Location Description” where the number of crimes is higher than expected? Come up with a table or visualization in which one can judge the frequency of crimes at each Location Description type.

=

* A measure was created for count of crimes based on ID column.
* We visualized it in a tabular form to get the count of crimes against the location description.
* On Location Description “Streets” the number of crimes is the highest with

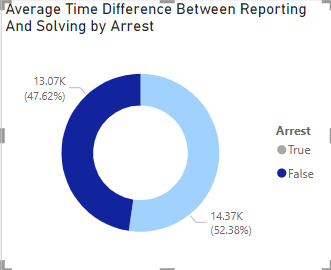
23,039 number of crimes reported.



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

=

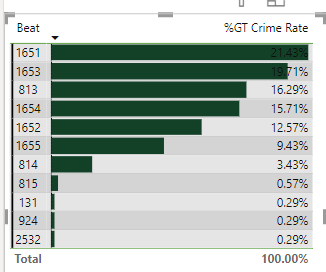
* I am considering a case solved if the Arrest has been done.
* I created a measure to calculate the average of time between date of reporting and updated on date.
* The average time for Arrested cases is 14,372.18 Minutes and for unsolved cases i.e. Non- Arrested Cases is 13,065.13 Minutes.



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

=

* The patrol areas that should be rewarded are specifically based on the criteria that their percentage of crime count is lesser than 40%.
* And their area of low crime count must be NON-SECURE AREAS.
* This shows these patrol officers are most active and vigilant as they can maintain low crime count in a non-secure areas.
* Patrol Areas are – Crime Count Percentage
  + 1651 – 21.43%
  + 1653 – 19.71%
  + 813 – 16.29%
  + 1654 – 15.71%
  + 1652 – 12.57%
  + 1655 – 9.43%
  + 814 – 3.43%
  + 815 – 0.57%
  + 131 – 0.29%
  + 924 – 0.29%
  + 2532 – 0.29%



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

=

* I created a calculated column for average time between reported date and updated on date.
* Add columns are created during the data loading phase in Power Query and it uses M language.
* Calculated columns are added after data has been loaded into the model and are part of the data model and it uses DAX language, which is designed for complex calculations and data analysis.

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

=

* I made one new measure to calculate the number of crimes based on these conditions.
* The count of crimes which are of type ‘theft’ and happened in ‘District 8’ is 1114.

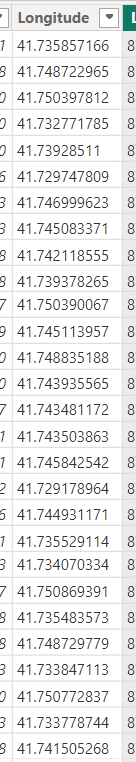


1. Using PowerBI can you separate the Longitude and Latitude from the Locations Column (Longitude, Latitude)? Which feature will you use?

=

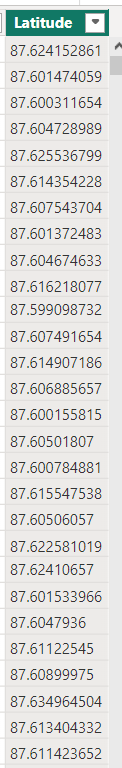
* We can separate the Longitude and Latitude from the Locations into two different columns using MID,LEFT AND RIGHT Functions.
* For Longitude Column I used a formula –

Longitude = MID('crimes\_data\_2022 - crimes\_data\_'[Location],FIND("(",'crimes\_data\_2022 - crimes\_data\_'[Location]) + 1,FIND(",",'crimes\_data\_2022 - crimes\_data\_'[Location]) - FIND("(",'crimes\_data\_2022 - crimes\_data\_'[Location]) - 1)



* For Latitude Column I used a formula –

Latitude = MID('crimes\_data\_2022 - crimes\_data\_'[Location],FIND("-",'crimes\_data\_2022 - crimes\_data\_'[Location]) +1,FIND(")",'crimes\_data\_2022 - crimes\_data\_'[Location]) - FIND("-",'crimes\_data\_2022 - crimes\_data\_'[Location]) - 1)



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?

=

* When we add a column in Power Query it formulates different formulas for different options in the add column feature.
* For Custom column feature it formulates the formula –

= Table.AddColumn(#"Previous Step", "NewColumnName", each [Column1] + [Column2])

* For Index column feature it formulates –

= Table.AddIndexColumn(#"Removed Columns", "Index", 0, 1, Int64.Type)

* For Conditional Column it formulates –

= Table.AddColumn(#"Previous Step", "NewColumnName", each if [Column1] > 10 then "High" else "Low")

* For date/time column it formulates –

= Table.AddColumn(#"Previous Step", "FutureDate", each Date.AddDays([DateColumn], 30))

M language is the formula language that executes in the Power Query Editor. Whenever a feature of power query editor is used the editor formulates a formula for its execution of that feature in the backend.

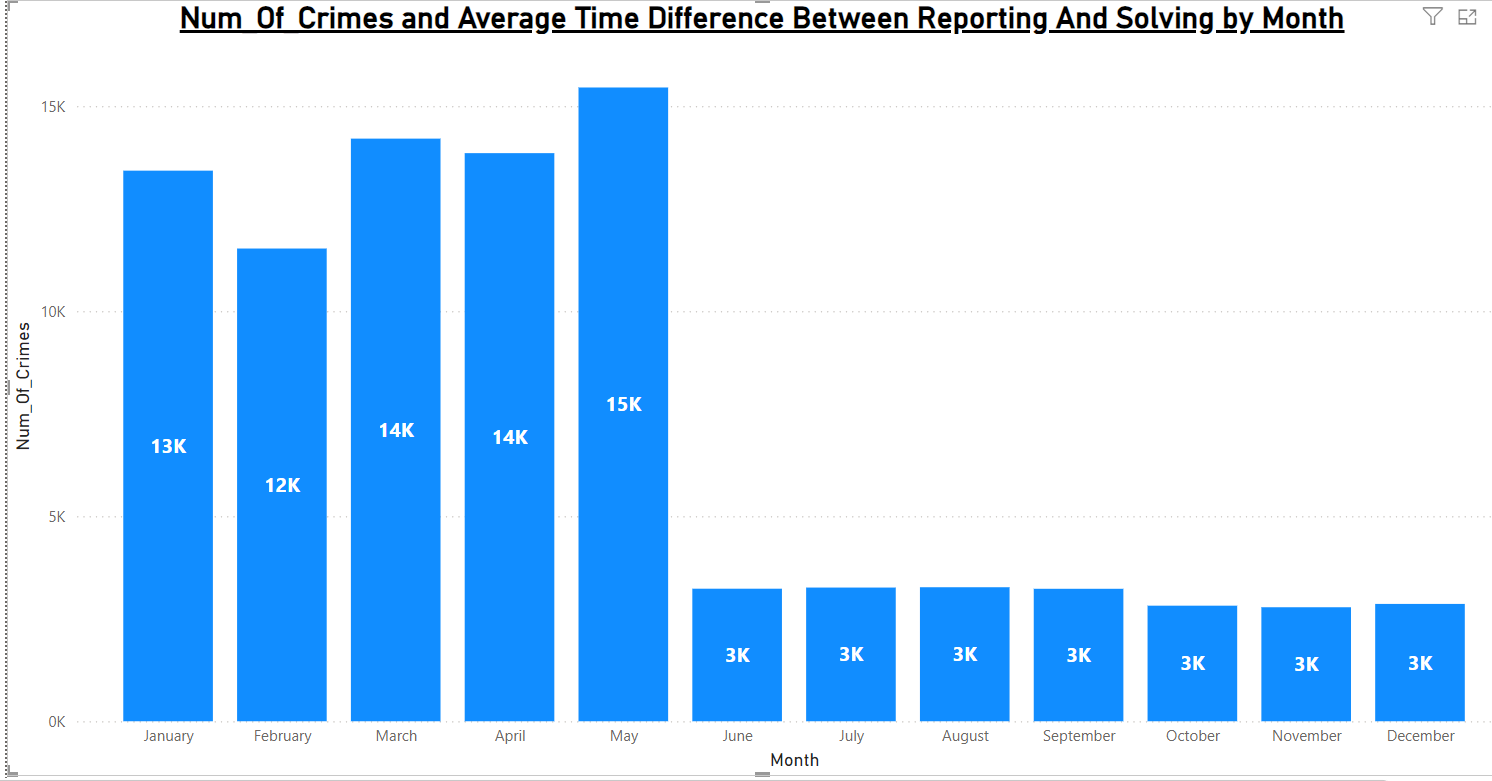
As for example when we use filter option in the editor it executes “FilteredRows = Table.SelectRows(PromotedHeaders, each [Column1] > 10)” this formulation in the back-end.

**Subjective Question:**

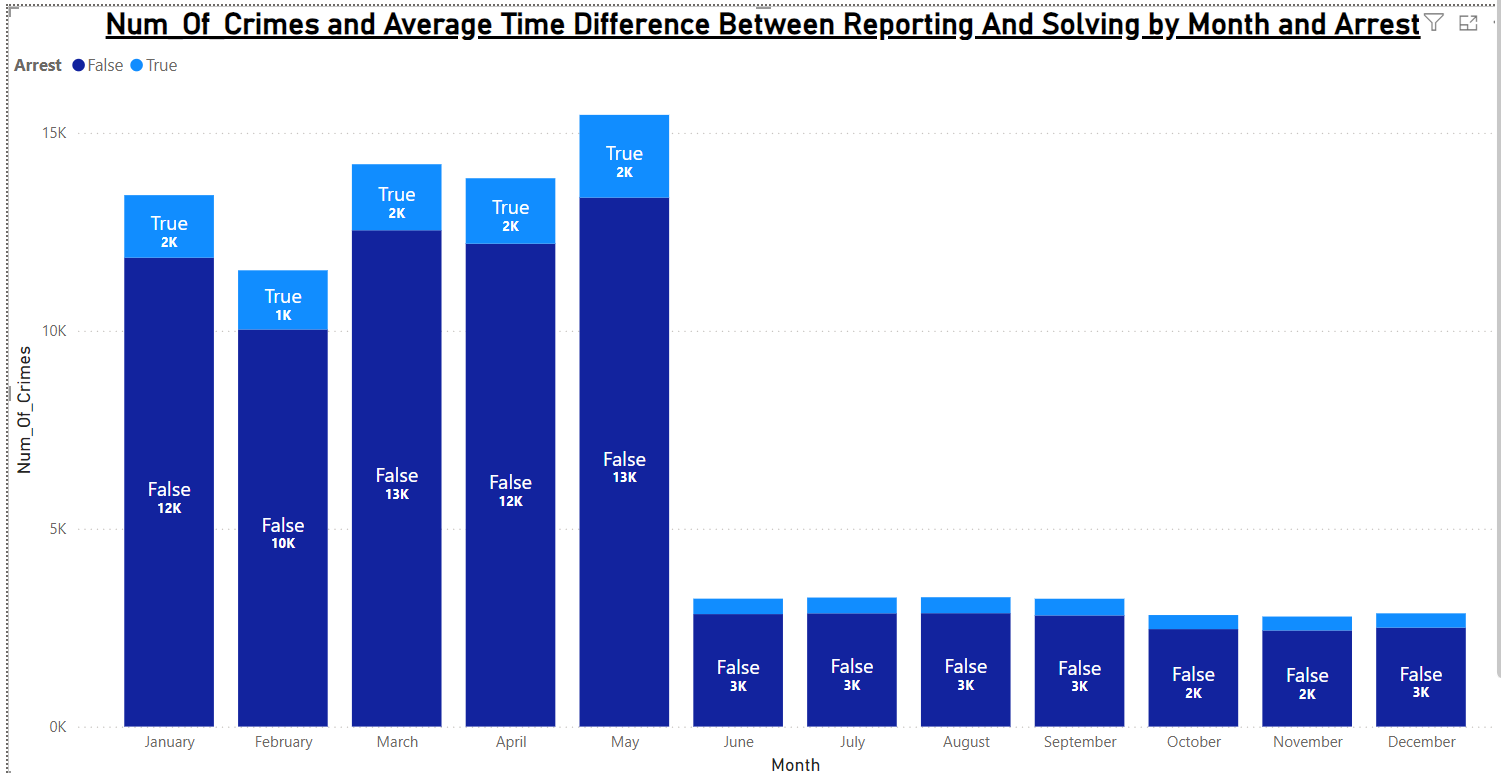
1. Is there any month-wise change in crime rates? If not, what could be the mistake in that operation?

=

* There is a month-wise change in crime rates in our current data.
* We have a significant decrease in crimes from May to June.
* The crime reported is lower from June onwards as compared to May and before.



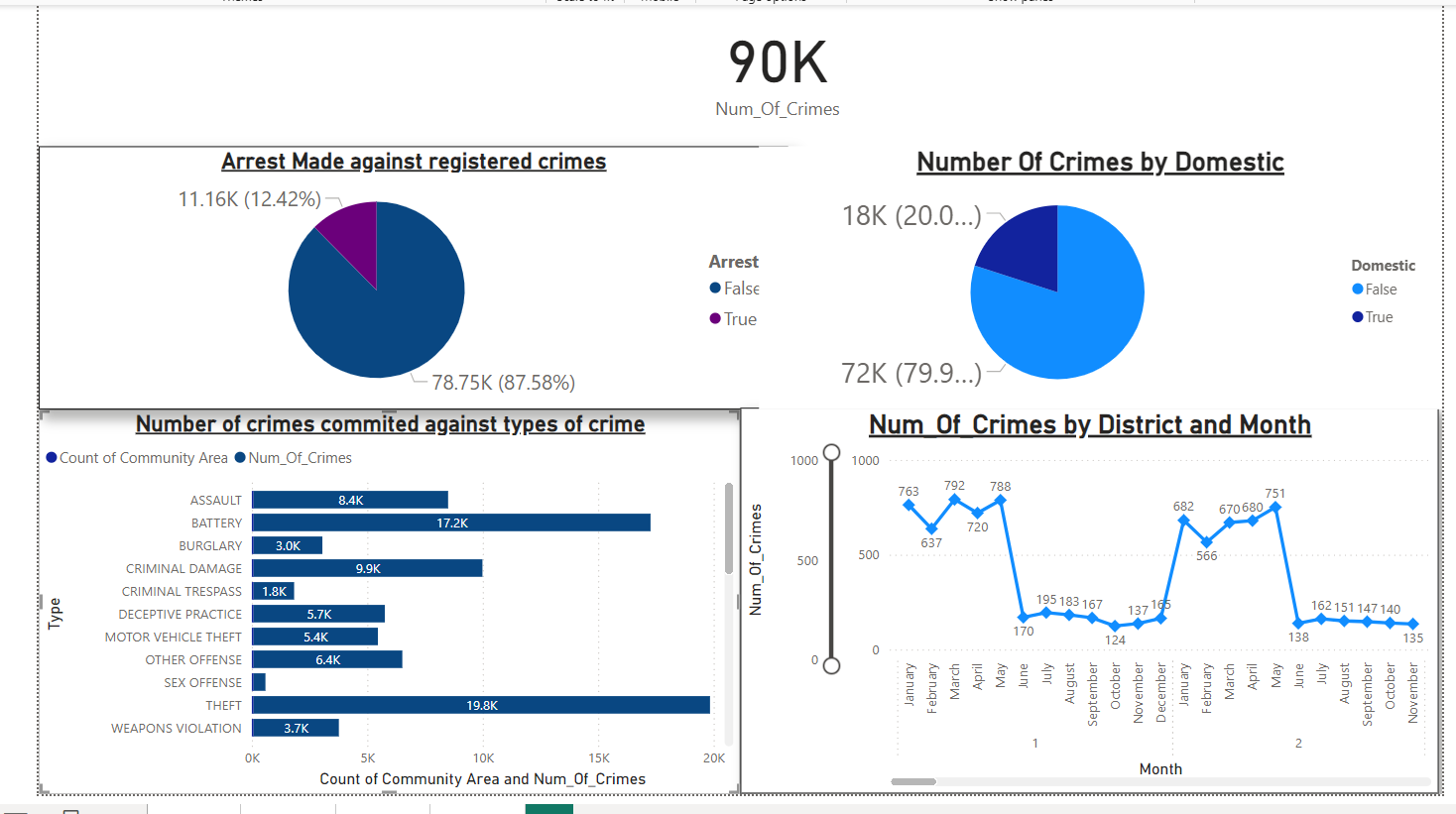
* If there’s no was no decrease in month-wise crime rates we could assume few possibilities.
* Such as, the average time between a reported crime and any update on it to be higher.
* The district or districts with high crime rate is not sufficiently patrolled to get the crime rate under control.
* The count of arrest made are not quite enough.
* Public’s vigilance towards crime is not quite enough.



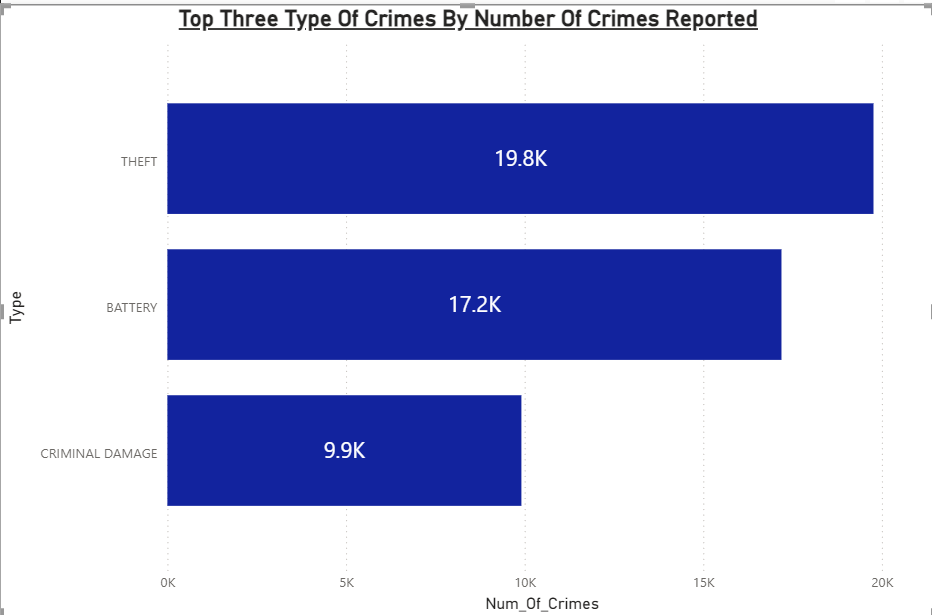
1. How can we reduce the no. of crimes, and which types of crime should we focus on to achieve improvement in the overall number of crimes?

=

* To reduce the crime rate-
  + We could increase patrol in the specific districts with high crime rate.
  + We can specify the location where the crime rate is high and focus on those parts primarily.
  + We could increase vigilance in public.
  + We could focus on increasing the arrest rate up to minimum of 50% according to the crime rate of that specific locality.



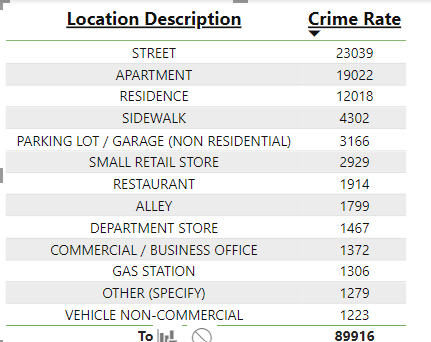
* The crimes that should be focused on to reduce the overall crime rate will be the most reported type of crime.
* The crimes that should be on focus are the top 3 most commited crimes.
* Those are –
  + Theft.
  + Battery.
  + Criminal Damage.



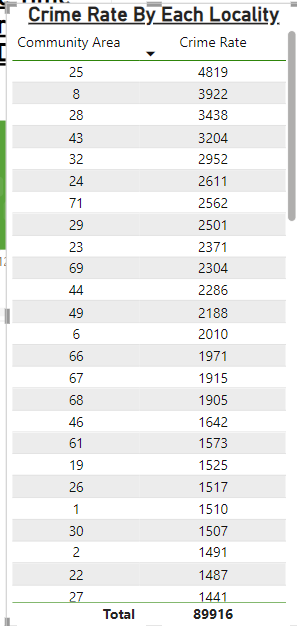
1. Which localities experience higher crime rates, and what measures can we ensure to reduce these numbers?

=

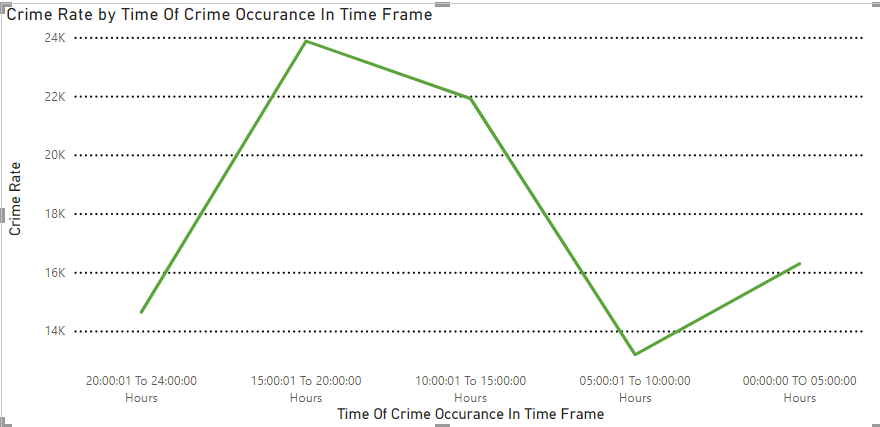
* We can view highest crime rates in the location of Streets and Second highest in Apartments.



* We can consider the community area to be the locality of each region.
* The community area 25 has the highest crime rate with 4819 crimes reported.



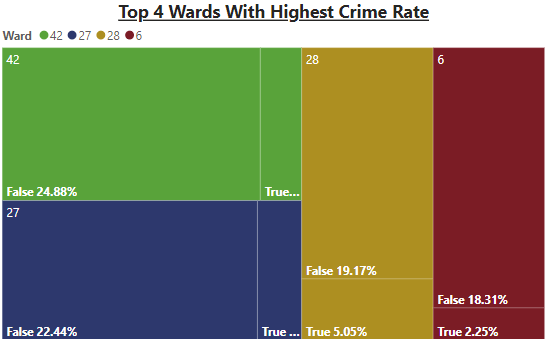
* Measures for reducing crime rate –
* We can introduce heavy patrol in the location description of high crime rate.
* We can increase the awareness in the public
* We can provide them the option of panic button which can be directly linked with the control room so that nearest patrol can be notified with their location.
* We can increase the patrols in the time of heavy crime occurrence.



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

=

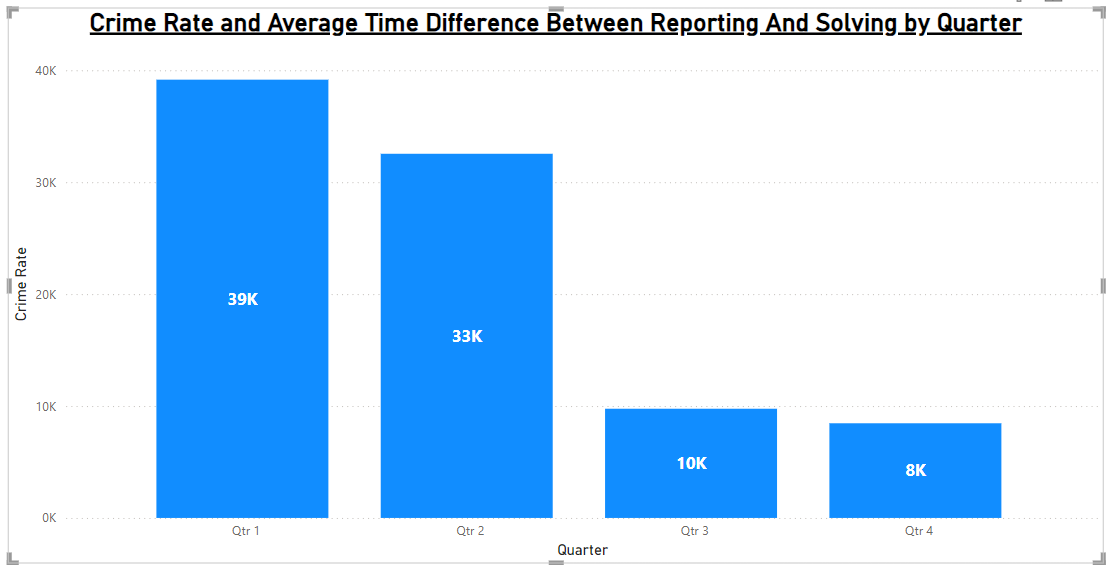
* We can narrow down the wards with the crime rate range in the highest.
* We considered the Top 4 wards with the highest crime rate to be the main focus for security improvements as they are affecting our overall crime rate.
* The wards in need of immediate security improvements are –
  + - * + Ward – 42
        + Ward - 28
        + Ward - 27
        + Ward – 6



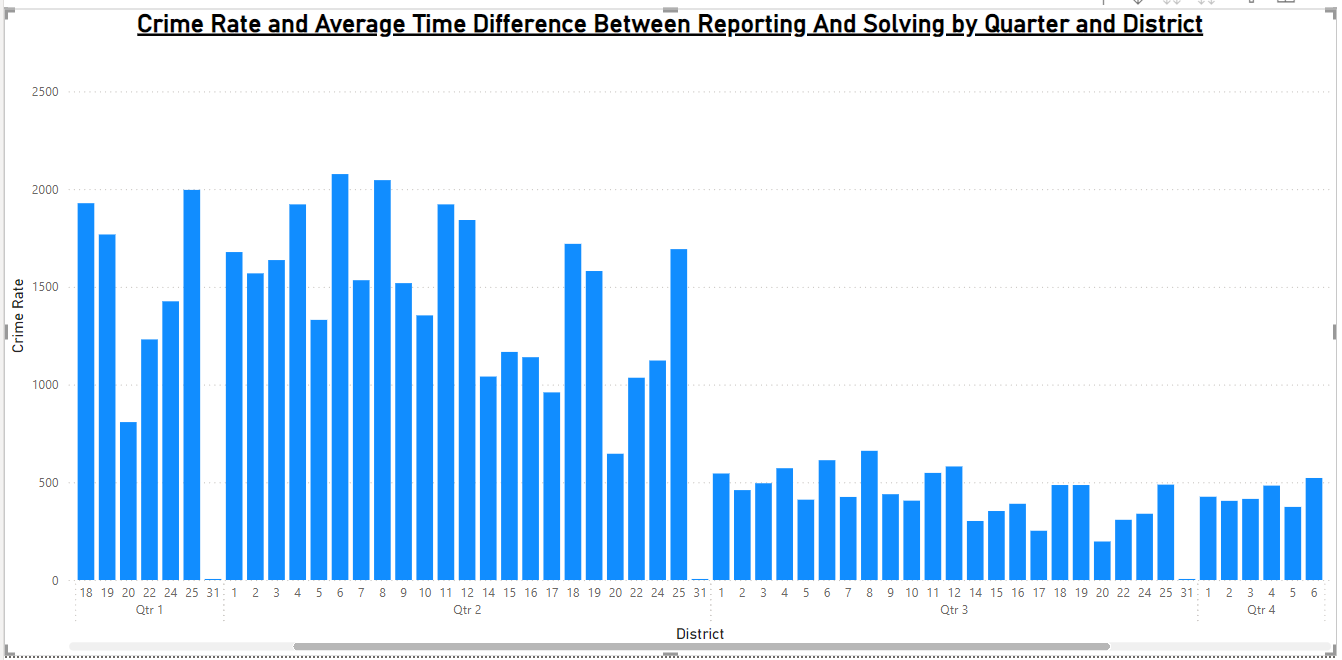
1. Crime Rate Trend Analysis: Monitor changes in crime rates over time to detect any discernible patterns or trends.

=

* For monitoring a crime trend we visualized a column chart in respect of quarters and months by district and crime rate.
* We can notice a trend of crime rate being high through out the months of Quarter 1 & 2.
* And a high drop of crime rate from Quarter 3 & 4.



* We can also notice that these type of pattern is followed through out every district.
* A high crime rate for Quarter 1 & 2 and a low crime rate for Quarter 3 & 4.



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)?

=

* For visualizing month-wise crime rate data we selected the table and took month from date hierarchy and our previous created measure of Total Crime Rate.
* For getting the previous month I created a dax formula measure to view the previous months Crime rate along with the current crime rate of that month.
* The formula used for that measure is - Previous Month Crimes =

VAR CurrentMonth = MAX('crimes\_data\_2022 - crimes\_data\_'[Date])

VAR PreviousMonthStart = DATE(YEAR(CurrentMonth), MONTH(CurrentMonth) - 1, 1)

VAR PreviousMonthEnd = EOMONTH(CurrentMonth, -1)

RETURN

CALCULATE(

[Total\_Crime\_Rate],

FILTER(

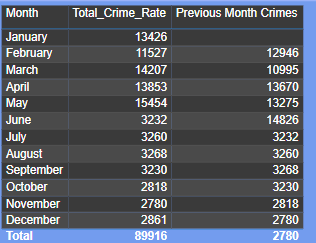
ALL('crimes\_data\_2022 - crimes\_data\_'),

'crimes\_data\_2022 - crimes\_data\_'[Date] >= PreviousMonthStart && 'crimes\_data\_2022 - crimes\_data\_'[Date] <= PreviousMonthEnd

)

)

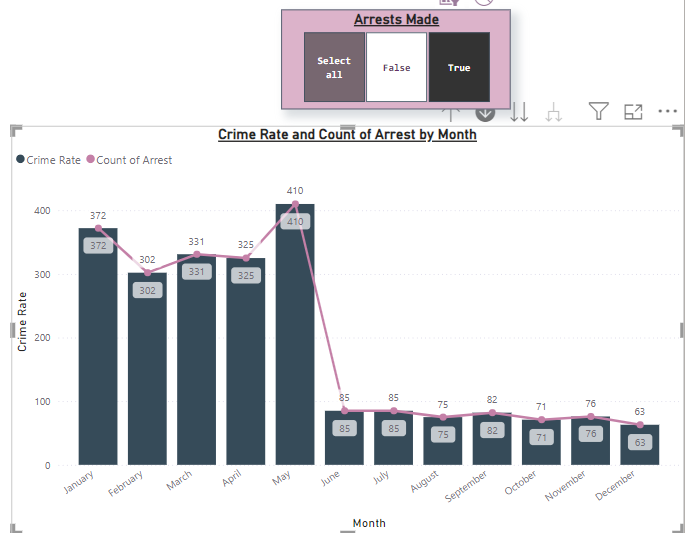
* This formula gave us the previous months sales along with the current months.



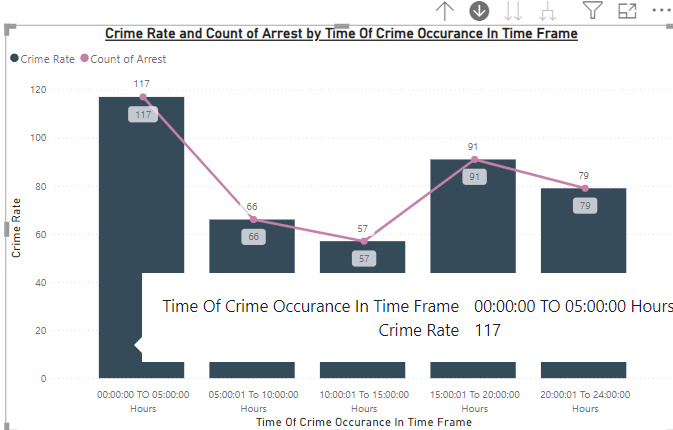
1. As per the previous reports, most domestic crimes do not result in arrest due to public hesitation and family pressure, is this trend also visible in our data?

=

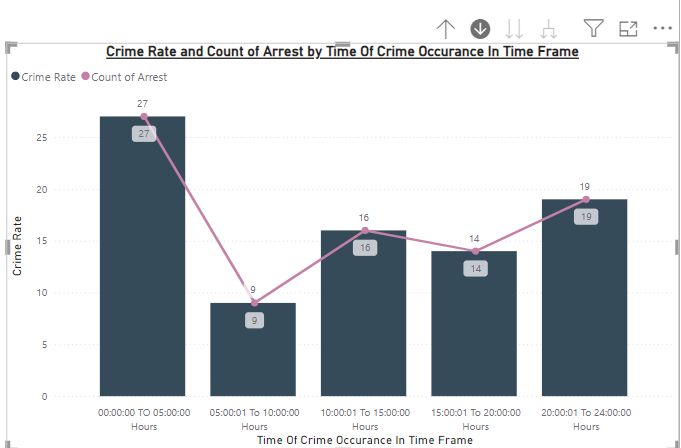
* We can visualize that there is a trend of arrest made for domestic crimes by months and time.
* Such as, there’s a trend of significant drop of arrests between January-May arrests and June – December arrests made.
* The number’s dropped from 410 to 85 i.e. there is a 79.5% drop in arrest made from May to June.



* I added a drill down on option on Time Of Occurrence’s to better understand the arrest made on domestic crimes are commited on which time frame for finding a trend.
* By clicking on the visualization drill down option we can access the time frame of the crime.
* For May –



* For June –

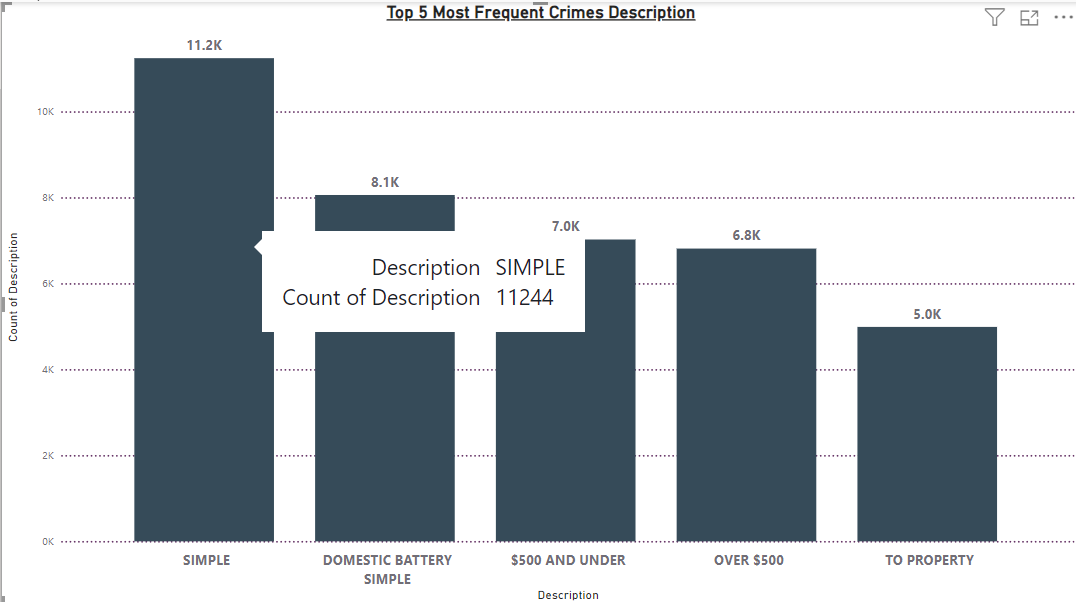


* We can view a trend of crime rate being high from 00:00:00 to 05:00:00 time frame.

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

=

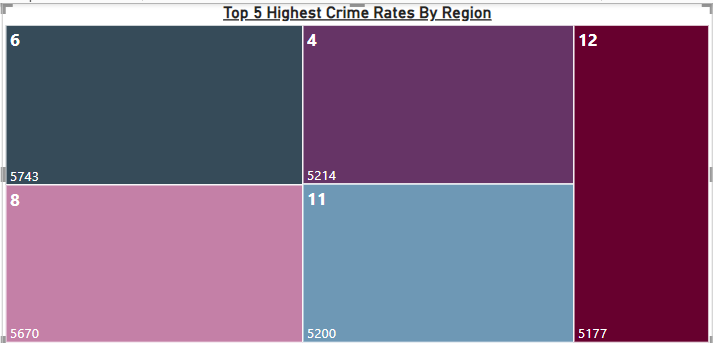
* We created a bar chart to find the top 5 most frequently occurring terms within the "Description".
* We created a count of description-by-description bar chart.
* And put one filter of Top N to Top 5 to filter out the Top 5 frequently occurring terms.
* The Top 5 Most Frequent Crimes Description are –
  + SIMPLE – 11244 Occurrences
  + DOMESTIC BATTERY SIMPLE – 8060 Occurrences
  + $500 AND UNDER – 7024 Occurrences
  + OVER $500 - 6813 Occurrences
  + TO PROPERTY – 4990 Occurrences



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

=

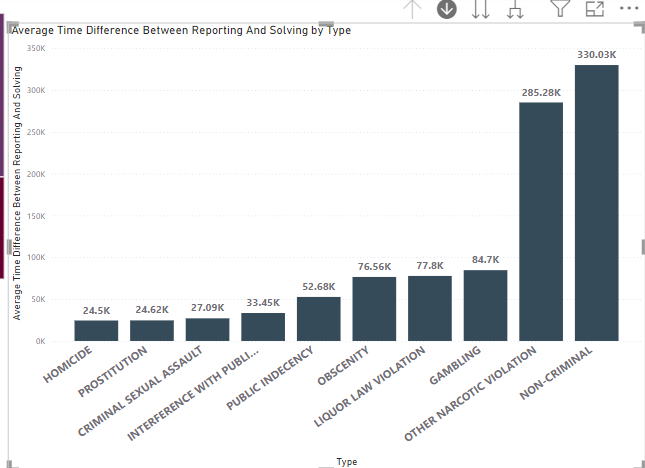
* I considered the districts as particular regions.
* We created a treemap visualization to better visualize the top 5 regions with highest crime rate.
* District with highest crime rate are-
  + - 1. District – 6
      2. District - 8
      3. District - 4
      4. District -11
      5. District -12



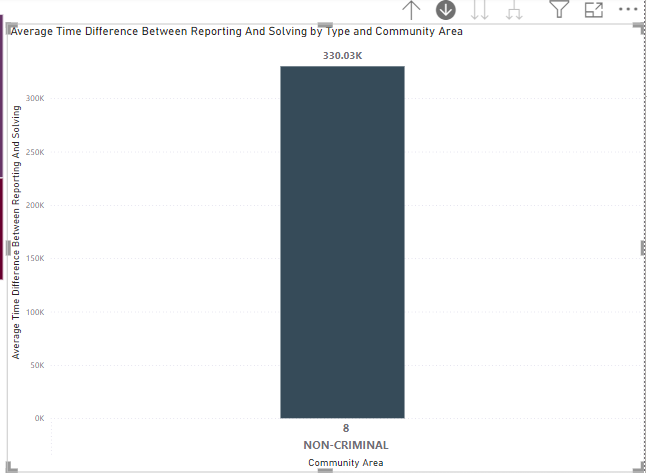
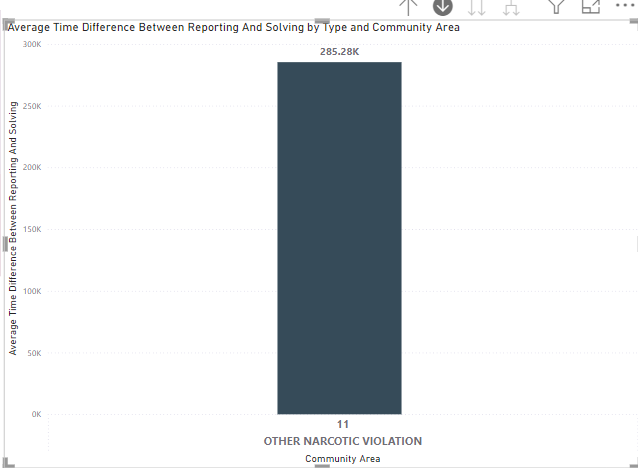
1. Is the solving time of cases also dependent upon the type and locality of crime?

=

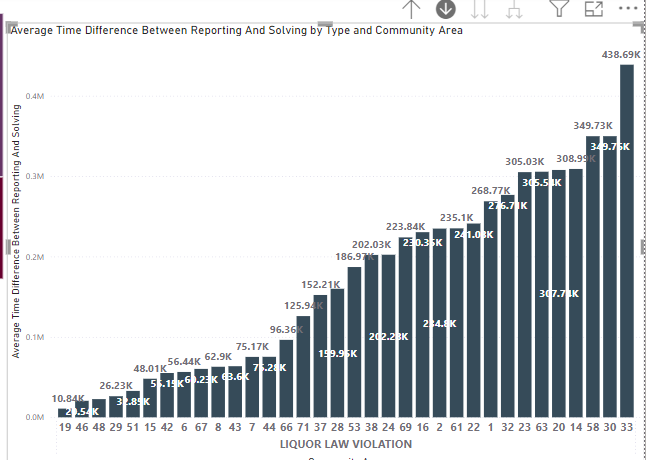
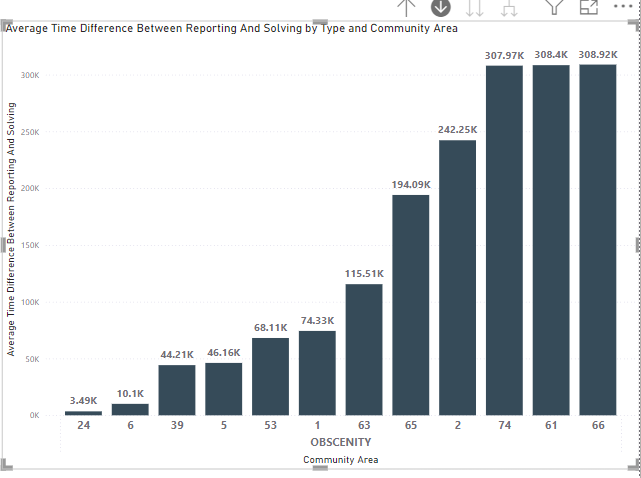
* The average solving time vary for different types of crime.



* The top 2 type of crimes NON-CRIMINAL and OTHER-NARCOTIC VIOLATION based on solving time are not dependent on locality i.e. the community area.

* But the crimes such as OBSCENITY and LIQUOR LAW VIOLATION are highly dependent on locality i.e. the community area.

* We can conclude that the average solving time of cases are highly dependent on the type and locality as per our current data.

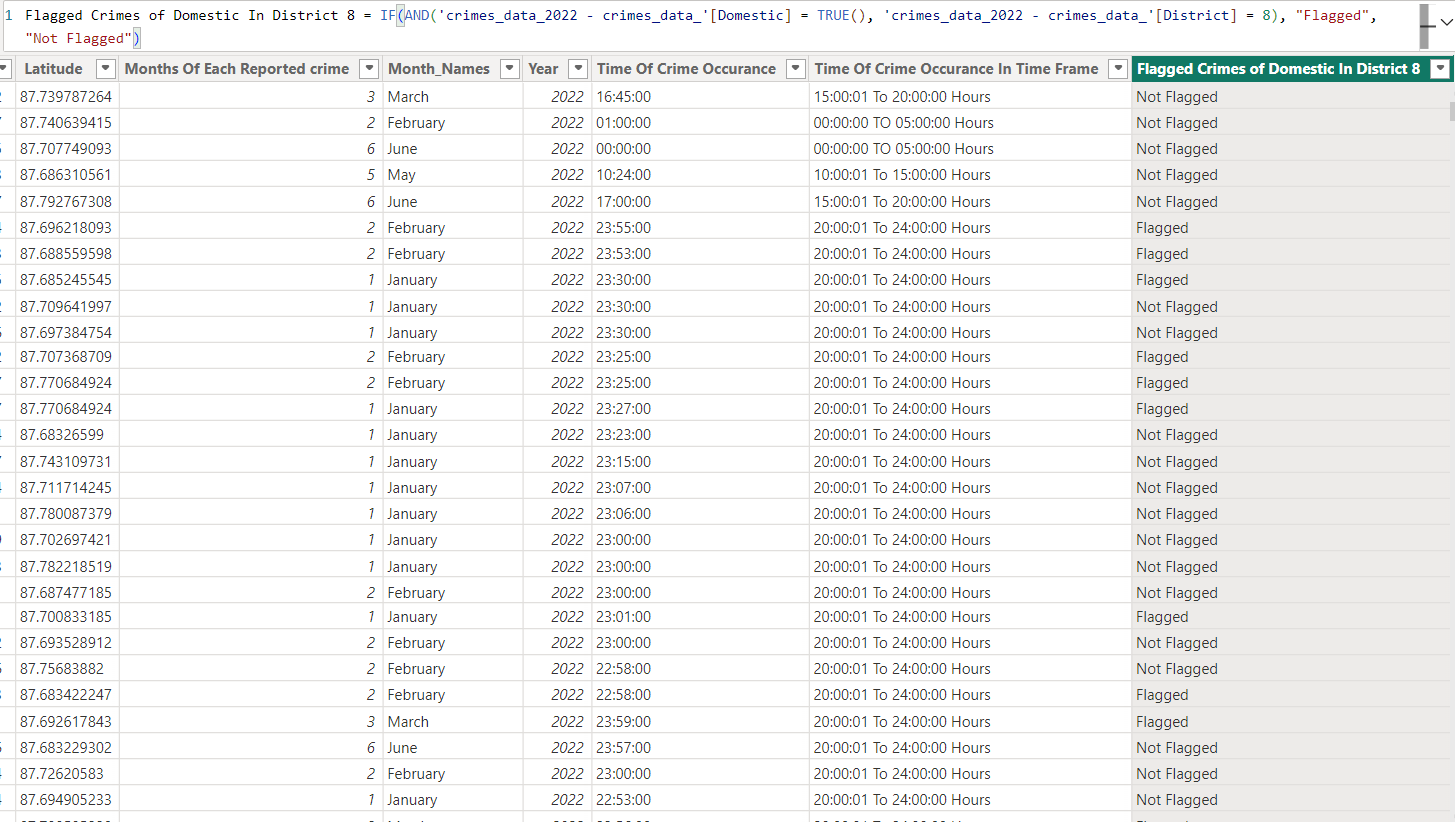
1. Create a calculated column to flag the domestic crimes that took place in District 8.

=

* As per the requirements we created a calculated column named “Flagged Crimes of Domestic In District 8”
* In that column we put text “Flagged” if that row of reported crime is domestic and from district 8.
* And Not Flagged if they doesn’t meet the above criteria.
* The formula used for this conditional column –

Flagged Crimes of Domestic In District 8 = IF(AND('crimes\_data\_2022 - crimes\_data\_'[Domestic] = TRUE(), 'crimes\_data\_2022 - crimes\_data\_'[District] = 8), "Flagged", "Not Flagged")

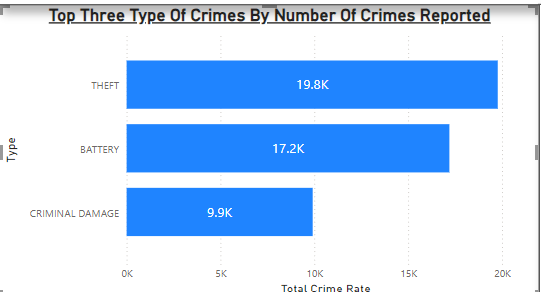
* And we cross checked by using filter if the function is working properly or not.
* We got required output in our calculated column.



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?

=

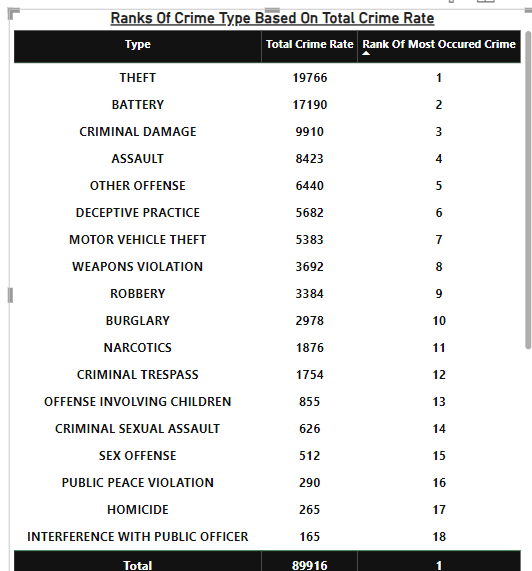
* As per my observation I can point out THEFT, BATTERY AND CRIMINAL DAMAGE as these crimes can lead to fatality.
* And the Crime Rate is the highest for these three types of crime.



* For ranking the crimes based on their count of occurrence’s we created a measure and used a RANX dax function.
* The formula used for ranking is –

Rank Of Most Occured Crime = RANKX(ALL('crimes\_data\_2022 - crimes\_data\_'[Type]),[Total Crime Rate],,DESC,Dense)

* We visualized it with the help of a table in report view.



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

=

* Understanding Of PowerBI gateway –
* The Power BI Gateway is essential for bridging the gap between on-premises data sources and the Power BI cloud service.
* It helps us set up schedule refresh of the data so that we can avoid manually updating it every time. Mostly efficient for Real-Time databases.
* We can handle multiple databases at the same time with the help of PowerBI gateway service.
* Use Cases –
* It’s ideal use case will be for connecting different data sources to a single dashboard and reports.
* We can use it to connect even local datbases which are on a specific system.
* It provides controlled access as well like the admin can decide which users will be able access which databases.

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

=

* I would have looked into every table to understand the data it consist of.
* Then I would have gone to power query editor to work on the null and inconsistence values or rows.
* I would have duplicates in the dataset.
* I would have gone to model view to see the connection and determine the relationships as well as assign some hierarchy and datatypes to the columns.
* Then I would have created a measure to calculate the crime rate at the very first.
* Then I would have created few visualization to catagories the Crime rate according to type, district, ward, community and also by arrest and domestic.
* Then I would have created some calculated columns for better understanding and catagorization of data such as time frame, avg time taken for case update.
* Then I would have also created a rank column for ranking the cases on average time taken for each solved cases.
* Then I would have checked the crime rate fluctuation based on dates and time frame for better understanding of the crime rate increase time frames.
* And analysed the data to find the highest crime on which month and time.
* And also which crime types are most occurring and at which time frame.
* Also which region is heavily contributing to thr crime rate increase and also which one is the least contributer.
* Then I would have created two dashboard to visualize data based on regions and types respectively.

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?

=

* The relation will be primary key to foreign key and fact table to dimension table.
* Our crime database will be the fact table and crime table’s district will be the foreign key.
* And districts-states table will be our dimension table and district\_id will be the primary key.
* With the help of that table we could have visualizes the district names based on crime rate which will provide us with more deep dive analysis.

**Report:**

The department has asked for a dashboard with three tabs:

1. Main Tab
2. Locality Tab
3. Type Exploration Tab

* **Using the Main tab in the report,** the government should be able to review the decrease in crime numbers and the improvements resulting from special police operations. This tab should include a date slicer and a filter for the primary type of crime.
* **Using the Locality Tab,** police management and the head of operations should be able to identify the most common types of crime and the localities/wards where these crimes occurred. This tab should include slicers forward, date, and primary crime type.
* **Using the Type Exploration Tab,** the Special Cops team aims to examine the total number of domestic crimes and, from those, how many arrests were made by the district. Additionally, they want to identify the crime type with the highest number of arrests. Essentially, this tab will provide metrics to summarize the department's actions and the number of tasks pending. The tab should include slicers for a month and the primary crime type.

**Make sure that all the visualizations look decent and are placed in a proper order. There are different POCs (Point Of Contact) for each tab, so make sure you involve all the metrics that POC may look at in that tab along with those mentioned in the tab description.**

**​​After making the report on the Desktop ensure that it is published on the PowerBI service and use the hosted link for submission of the dashboard and mentioning on the resume.**