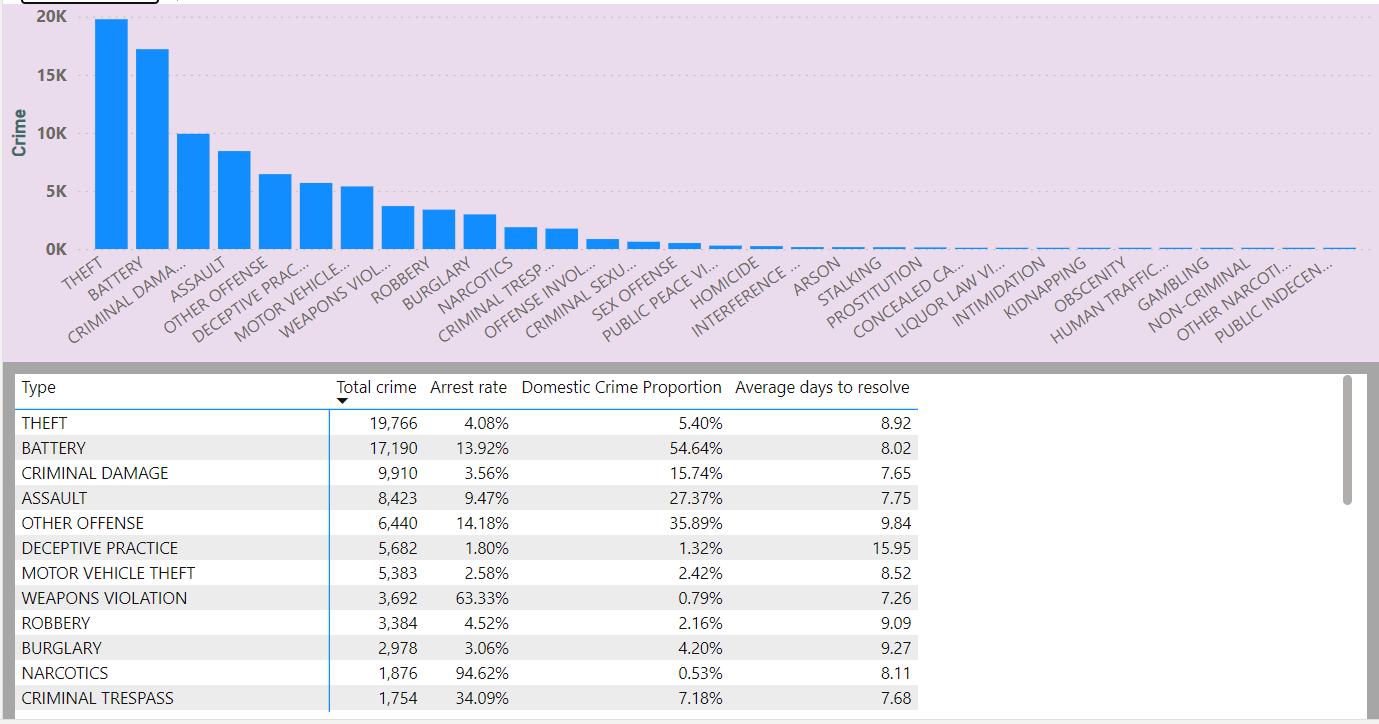
Objective Questions:

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

**Ans :-**

1. **Handling Null Values:** In order to prioritize accuracy in spatial analysis and visualizations, I addressed null values, particularly for the X coordinate. I made a conscious decision to keep blank values in the 'Location Description' and 'Ward' columns, understanding that these fields may be legitimately blank due to lack of information.
2. **Creating New Columns :** 
   1. **Latitude and Longitude:**
      1. Created using the Location column through the "Add Column by Example" feature.
   2. **Day in week:** 
      1. Created using dax formula -> Day in week = FORMAT('crimes data 2022'[Date],"dddd")
   3. **Moment of Case Register:**
      1. Extracted the last 2 letters from the date.
   4. **Arrest Binary:**
      1. Assigned 1 for True and 0 for False from Arrest Column.
      2. Using the formula Arrest Binary = IF('crimes data 2022'[Arrest], 1, 0).
   5. **Resolve Case Time:**
      1. Calculated the date difference between Date and Updated.
      2. Using the formula Resolve Case Time = DATEDIFF('crimes data 2022'[Date], 'crimes data 2022'[Updated On], DAY).
   6. **Month:**
      1. Extracted the month from the Date column using the formula Month = FORMAT('crimes data 2022'[Date], "MMMM").
3. **Transforming Data Types:**
   1. Date :- Extracted Date from the table column as give as date datatype
   2. Text Datatype :- Beat, District, Ward, Community Area, FBI Code to Text data type.
4. **Crime Type Analysis: Assess the frequency of each crime type to identify the most prevalent crimes occurring in the area.**

**Ans :-** To understand the distribution of crime types, I calculated a "Total Crime" measure and visualized it with a clustered column chart. This analysis showed Theft as the most prevalent crime, accounting for 19,766 reported incidents.



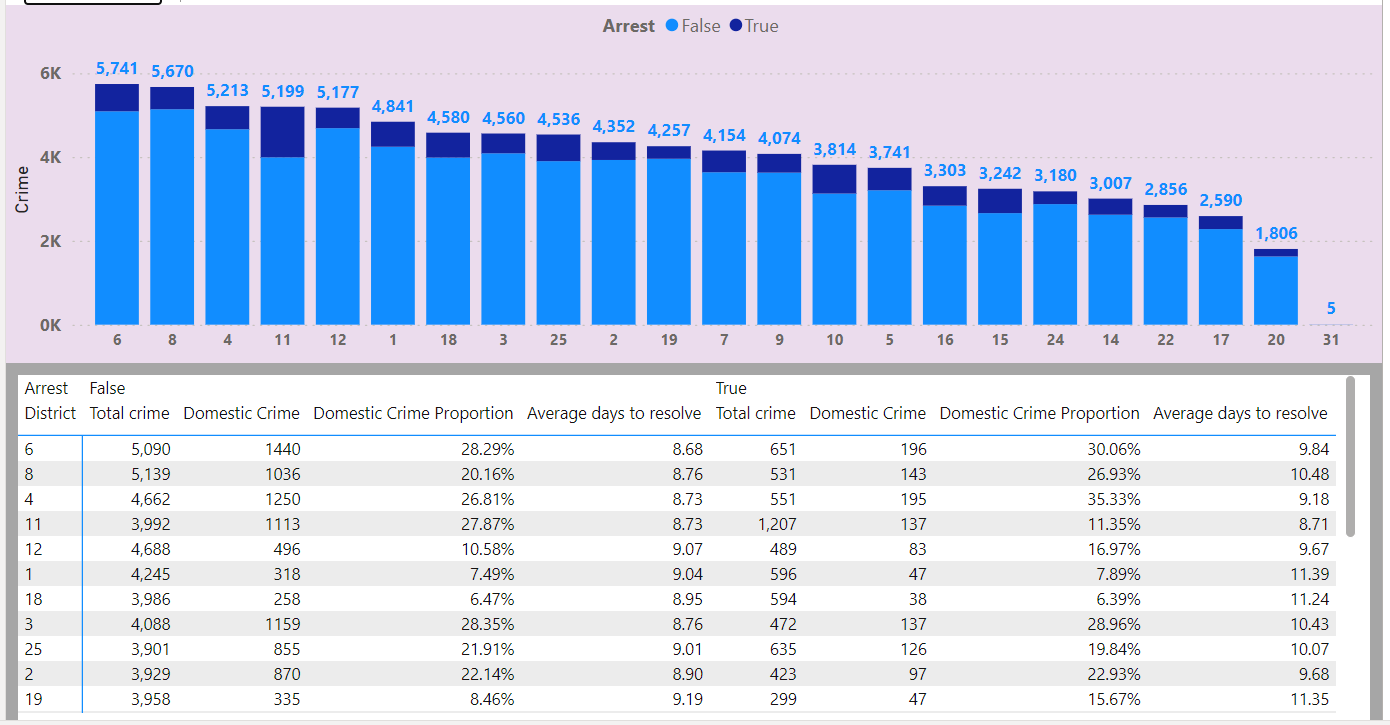
1. **Arrest Rate Evaluation: Analyze the percentage of reported incidents that have resulted in an arrest to gauge law enforcement effectiveness.**

**Ans :-** I Calculated the arrest rate, which are 12.42%, using the formula Arrest Rate = AVERAGE ('crimes data 2022'[Arrest binary]).



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 :-** Unveiling patterns in crime location, I utilized a stacked column chart to depict the crime volume in each district. The visualization clearly identified District 6 as the most crime-ridden area, with a concerningly high number of 5,741 reported cases.



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

**Ans :-** The Followings are the categorical attributes in the data.

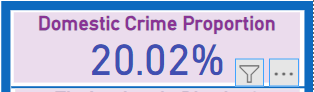
1. Case Number.
2. Block.
3. IUCR.
4. Type.
5. Location Description.
6. Arrest.
7. Domestic.
8. Beat.
9. District.
10. Ward.
11. Community Area.
12. FBI Code.
13. Year.
14. **Were there any Null values in the data, if there were how did you handle them? What is the ideal way to handle Null values?**

**Ans :-** I found 944 missing location points (X co-ordinate and Y co-ordinate), which is only about 1% of the data. Since they're such a small number, I decided to remove them to keep my data clean.

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

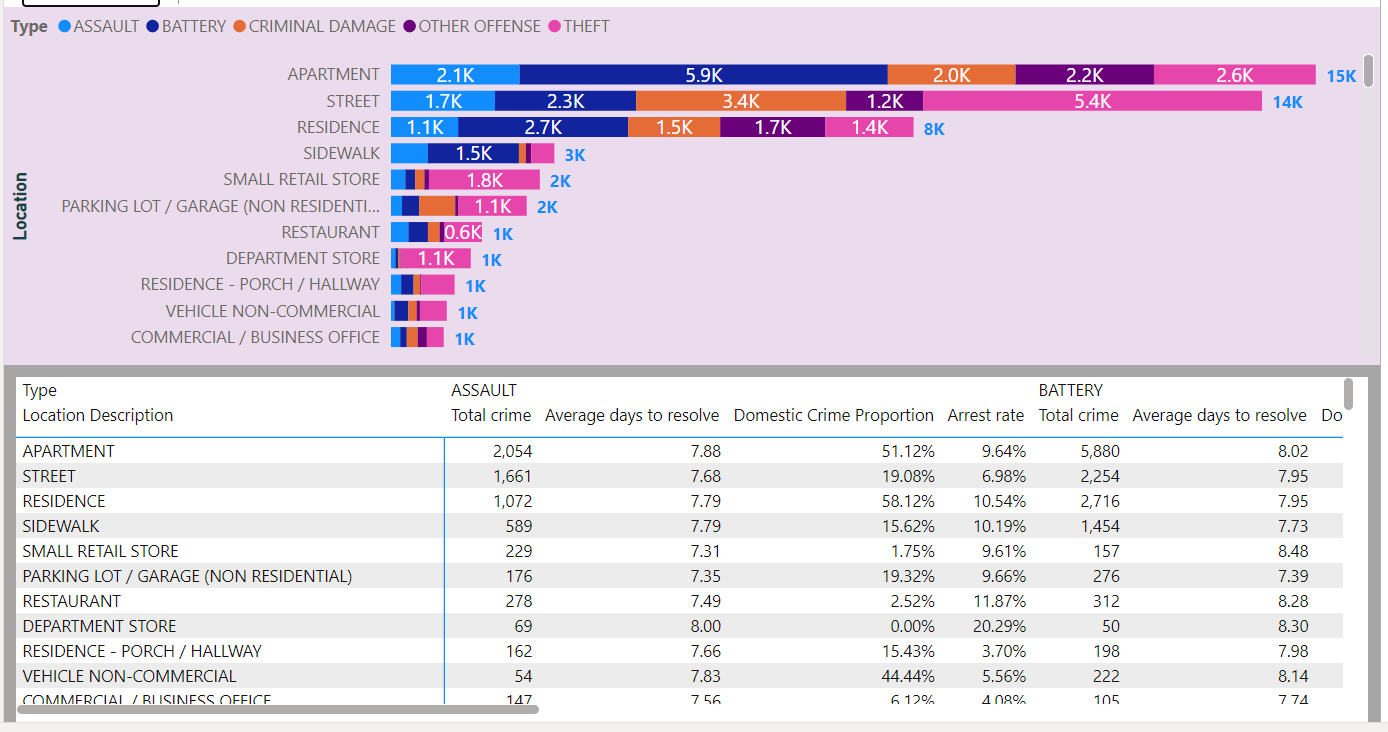
**Ans :-** My analysis revealed that domestic violence makes up 20.02% of all crimes in Chicago. To display this finding, I added a card on the "Main" tab showing the "Domestic Crime Proportion." This number is calculated using a formula that divides the total number of domestic crimes by the total number of crimes (where "Domestic binary" indicates whether a crime is domestic-related).

Domestic Crime Proportion = DIVIDE(SUM('crimes data 2022'[Domestic binary]), COUNT('crimes data 2022'[Domestic binary]))



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 :-** **Apartment** location is the highest crime frequency i.e 14.8K incidents.



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

**Ans :-** Using the formula :

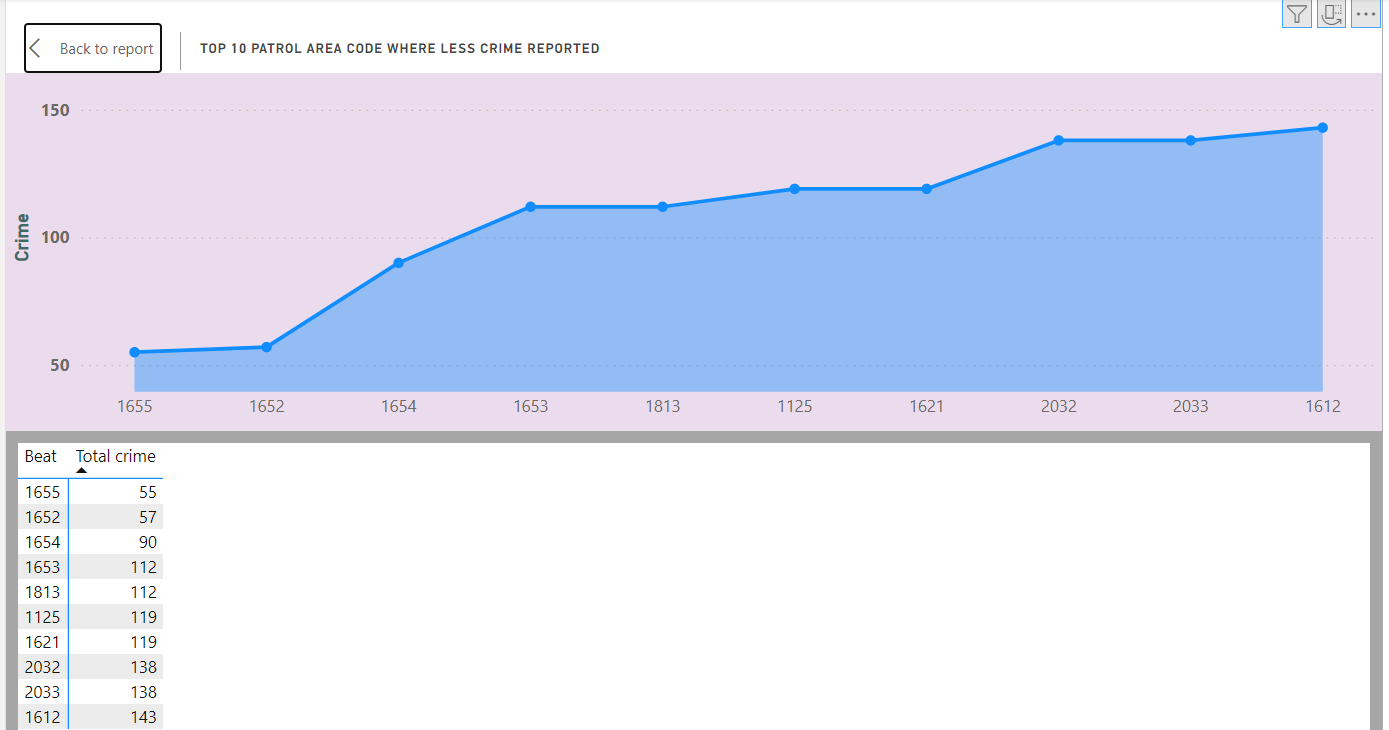
AVERAGE('crimes data 2022'[Resolving case time])

I found the average time to solve a crime is 9.03 days. This calculation, based on the "Resolving case time" data, shows us the usual amount of time it takes to close cases.



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

**Ans :-** To analyze crime distribution across beats, I used an area chart. It revealed Beat 1655 as a clear outlier, with a mere 55 reported crimes. This indicates that this police-patrolled area experiences substantially less criminal activity compared to other beats.



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

**Ans :-** **My analysis involved creating new data points to enrich the dataset. Here's what I did:**

* **Moment of Case Registration:** I extracted the last two characters from the date using the "Extract Last 2 Characters" function. This created a new column revealing the time of day a case was registered, offering insights into potential registration patterns.
* **Longitude and Latitude:** I used Power BI's "Column from Example" feature to automatically generate these columns based on sample data I provided. This helped pinpoint the location of each crime incident.

**In addition to these, I also created 8 custom formulas (called calculated columns) using DAX. These formulas performed calculations, comparisons, and text manipulations to generate even more data points.** This extra information helped me analyze the data more thoroughly and uncover deeper trends.

**Here's the key difference between calculated columns and adding columns:**

* **Calculated Columns:** Created within Power BI using formulas, these columns are permanent additions to the data model. They can be used throughout your analysis for various purposes.
* **Add Columns (Custom Columns):** Used in Power Query Editor, this feature lets you create new columns by transforming existing data. These columns are temporary and not stored in the model, but are helpful for cleaning and preparing your data.

**In short, both methods create new data points, but calculated columns are permanent for analysis, while custom columns are temporary for data manipulation.**

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 :-** To find out how many thefts happened in District 8, I used a formula to filter the data. The formula looked at crimes labelled as "theft" and only included those that occurred in "District 8". This calculation shows there were 1143 reported thefts in District 8.

Theft crime in District 8 = CALCULATE([Total crime], AND('crimes data 2022'[Type]= "theft",'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?**

**Ans :-** In Power Query Editor, I used the "Add Column with Example" feature to separate values that were originally combined with commas. This made it easier to work with the data and organize it for clearer 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 :-** **Think of M language as the secret code Power Query uses to follow your instructions.**

* **Adding Columns by Splitting:**
  + Let's say you want to split a "Location" column based on commas. The M code would use the Table.SplitColumn function to understand how many new columns to create and what to name them. It would also reference the comma (",") as the separator.
* **Adding Columns by Extracting Text:**
  + If you want a new column showing the last two characters of a "Date" column (think AM/PM for registration time), the M code would use Table.AddColumn to create a new column and Text.End to grab those last two characters. It would also specify "text" as the data type for the new column.

**M language gives Power Query the precise instructions to follow when you use the "Add Column" feature.** It's a powerful tool for data manipulation, but you don't necessarily need to memorize the code itself. Just focus on the clear steps you want Power Query to take, and the M language will handle the technical side!

Subjective Question:

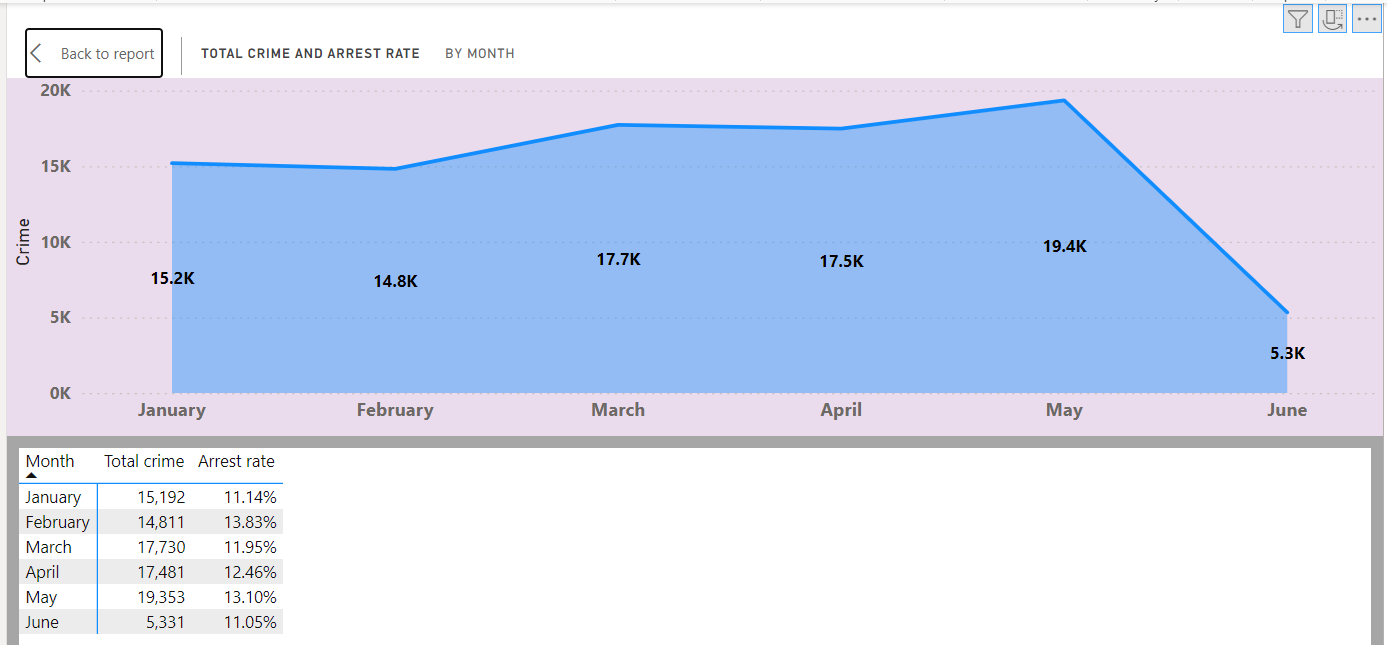
1. **Is there any month-wise change in crime rates? If not, what could be the mistake in that operation?**

**Ans :-** **Looking at crime rates in Chicago from January to June 2022 (with data only up to June 9th) reveals an interesting trend.** There's a gradual rise in crimes until May, followed by a sudden drop in June. This decrease might be due to the limited data for June.

**May stands out as the month with the highest number of reported crimes (19,353).** Understanding why crime spiked in May is crucial.

**On a brighter note, February saw the highest arrest rate (13.83%).** However, it's important to see what led to this increase in arrests during that specific month.

**By focusing on what caused the May crime surge, we can develop strategies to reduce overall crime in Chicago.**



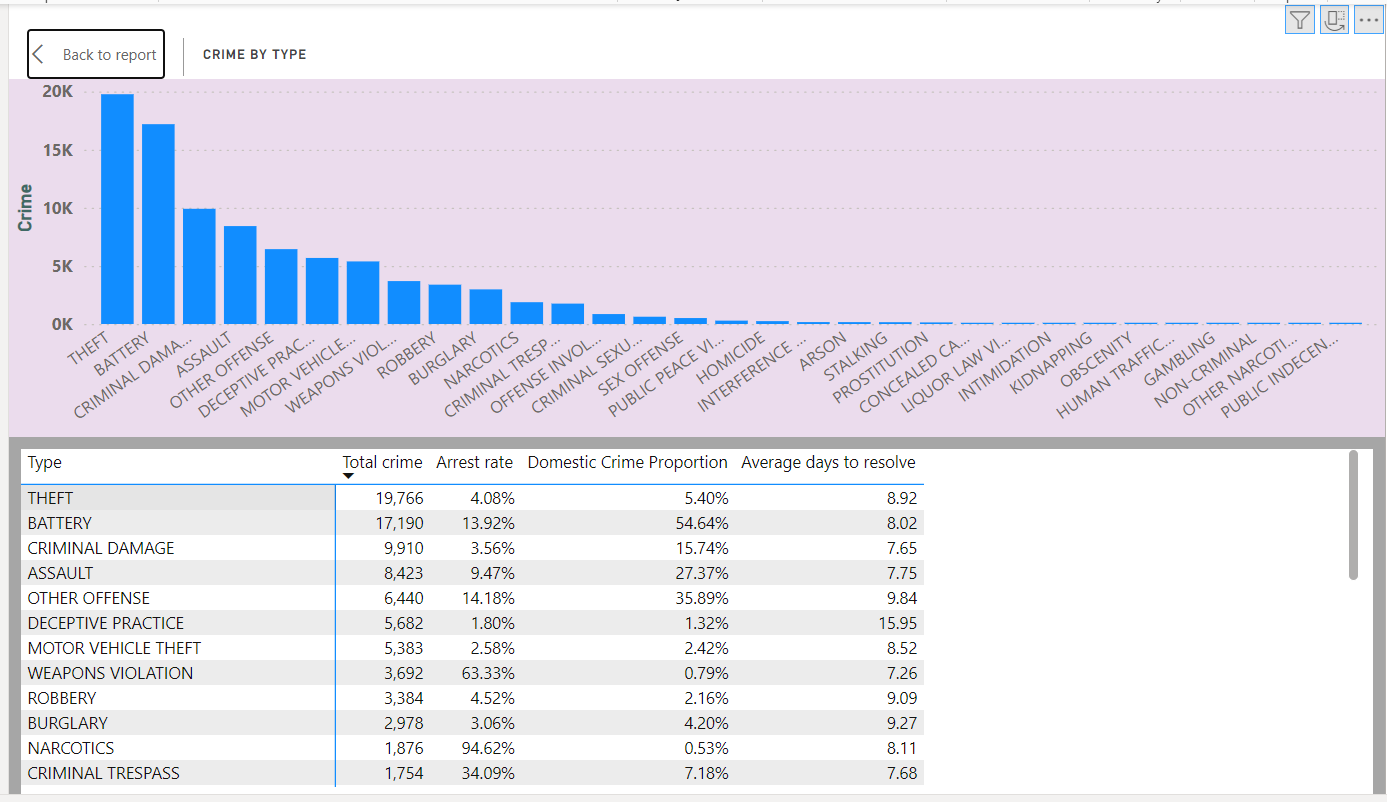
1. **How can we reduce the no. of crimes, and which types of crime should we focus on to achieve improvement in the overall number of crimes?**

**Ans :-** **Our analysis of crime data shows theft and battery are the biggest problems.** There were a whopping 19,776 thefts and 17,190 batteries. Interestingly, it's much harder to catch thieves (arrest rate of 4%) compared to battery offenders (arrest rate of 14%). Also, domestic violence is a major factor - over half (54%) of battery cases are domestic violence related, while only 5% of thefts are.

**To tackle crime, we need a smart plan that considers everything we learned.** Here are some key areas to focus on:

* **Education & Awareness:**
  + Teach people about the laws and what happens if they're broken.
  + Encourage people to report suspicious activity to the police.
* **Stopping Domestic Violence:**
  + Figure out why domestic violence leads to battery so often.
  + Work with social services to help people in abusive situations.
* **Police:**
  + Put more officers in high-crime areas for theft and battery.
  + Build trust with the community through better collaboration (community policing).
* **Surveillance & Technology:**
  + Use more cameras and other tech to catch thieves in high-theft areas.
  + Use data analysis to predict where crimes might happen.
* **Law & Enforcement:**
  + Make sure laws are strong and enforced quickly to deter crime.
  + Increase arrest rates to discourage criminals.

**The data shows crime reduction is complex, but with a multi-pronged approach, we can make Chicago safer.** We need to focus on high-crime areas, address domestic violence, and strengthen law enforcement and policing strategies.



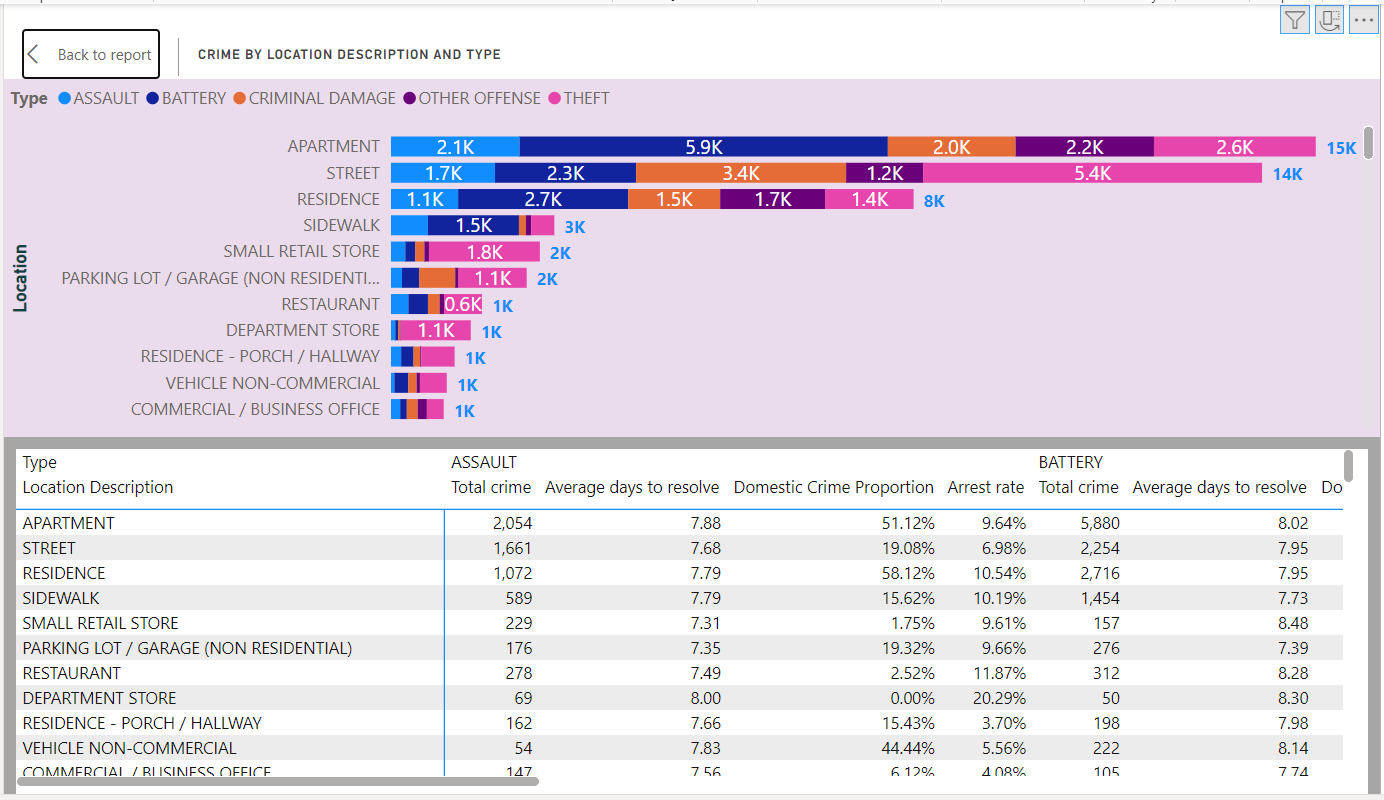
1. **Which localities experience higher crime rates, and what measures can we ensure to reduce these numbers?**

**Ans :-** **Our data shows crime is highest on Apartment Street (14,768 incidents), Residence Location (13,910 incidents), and another area (8,345 incidents).** Thefts and batteries are the most common crimes in these areas, and the arrest rates are lower than we'd like, especially for domestic violence.

**To bring down crime in these neighborhoods, we can try a few things:**

* **More Police & Cameras:**
  + Increase patrols and surveillance, especially on Apartment Street and Residence Location.
  + Use more security cameras to watch for and deter crime.
* **Working with Neighbors:**
  + Start neighborhood watch programs where residents look out for each other.
  + Build a stronger sense of community to discourage crime through everyone being aware.
* **Safer Streets:**
  + Improve street lighting to make it harder for criminals to hide at night.
  + Design streets and buildings in a way that discourages crime (think good lighting in public areas).
* **Helping People:**
  + Invest in education and job training to give people opportunities and keep them out of trouble.
  + Make it easier for people to get help with mental health and addiction issues.
* **Keeping Kids Safe:**
  + Offer programs to keep young people out of crime and on the right track.
  + Push for stricter gun control laws to make it harder for criminals to get guns.

**The key is to combine police work with social programs to tackle crime from all angles.** We need to address what's happening now and also fix the root causes of crime to make these areas safer in the long run.



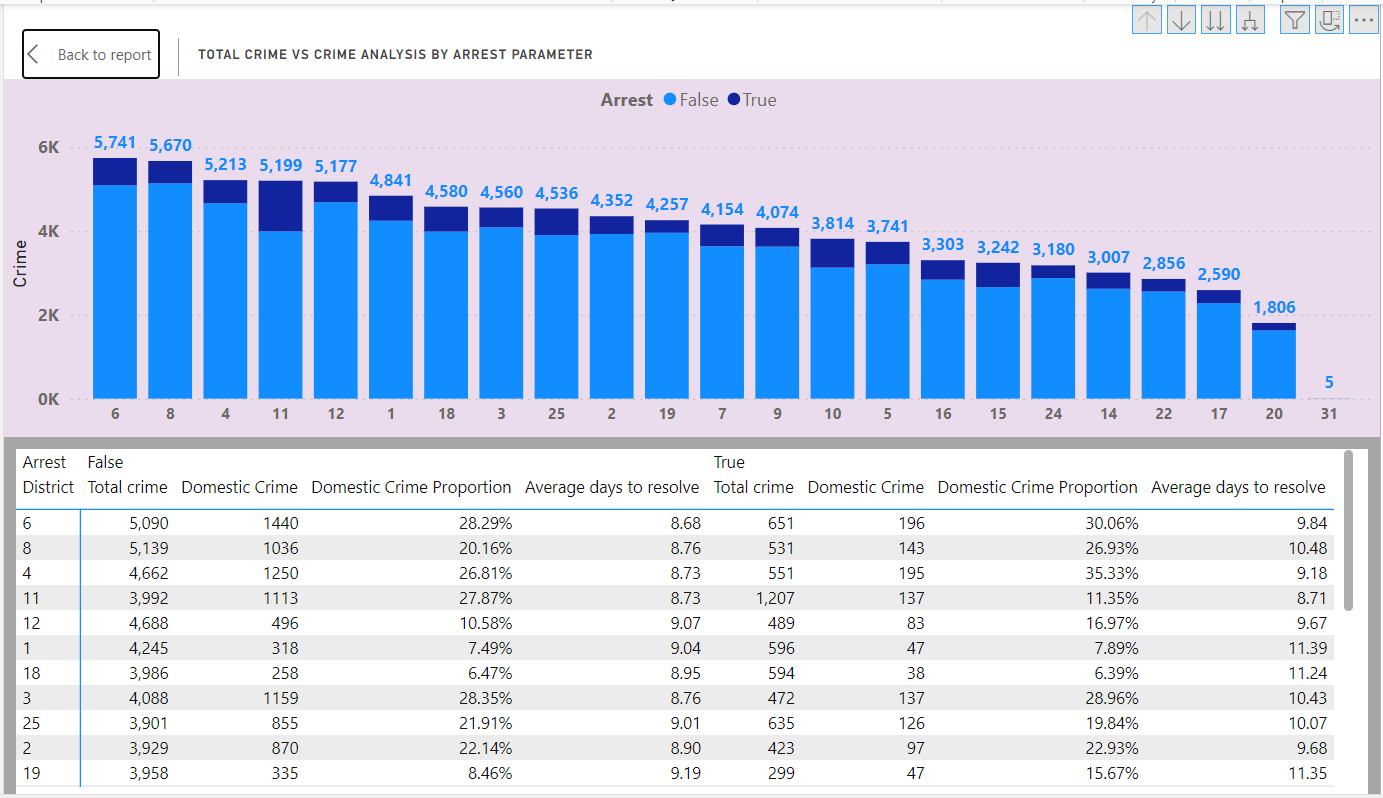
1. **Can you suggest wards where security improvements should be made to reduce crime?**

**Ans :-** **The data shows some areas are struggling more with crime than others.** Wards 42, 27, 28, 6, and 24 all have over 3,000 reported crimes. Ward 7 also has a higher rate of domestic violence.

**To make these areas safer, we can try a combination of strategies:**

* **More Police & Community Policing:**
  + Put more officers on patrol in the high-crime wards (42, 27, 28, 6, 24).
  + Work with residents to build trust and encourage them to report crime (community policing).
* **Cameras & Lighting:**
  + Install security cameras in key areas to deter crime and help investigations.
  + Improve streetlights to make it harder for criminals to hide at night.
* **Helping Young People:**
  + Invest in programs and activities to keep young people out of trouble.
* **Mental Health & Addiction Services:**
  + Make it easier for people to get help with mental health and addiction issues.
* **Jobs & Community Development:**
  + Create more jobs and improve the overall community to reduce poverty-related crime.
* **Security Measures for Different Areas:**
  + Implement targeted security measures in different places, like public transport stops, neighborhoods, businesses, parks, schools, and high-crime zones.
* **Fight Cybercrime:**
  + Invest in cybersecurity to protect against online crime and educate people about staying safe online.

**By using a mix of these approaches, we can tackle crime in different ways and make these wards safer for everyone.**



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

**Ans :-** **Our crime data for 2022 shows some interesting patterns.** Crime peaked in May, and Tuesdays and Saturdays saw the most crime each week.

**It's important to remember that this is just one year of data.** To see clearer trends, we should look at data from multiple years. This would help us understand how crime rates change over time and if there are any seasonal patterns.

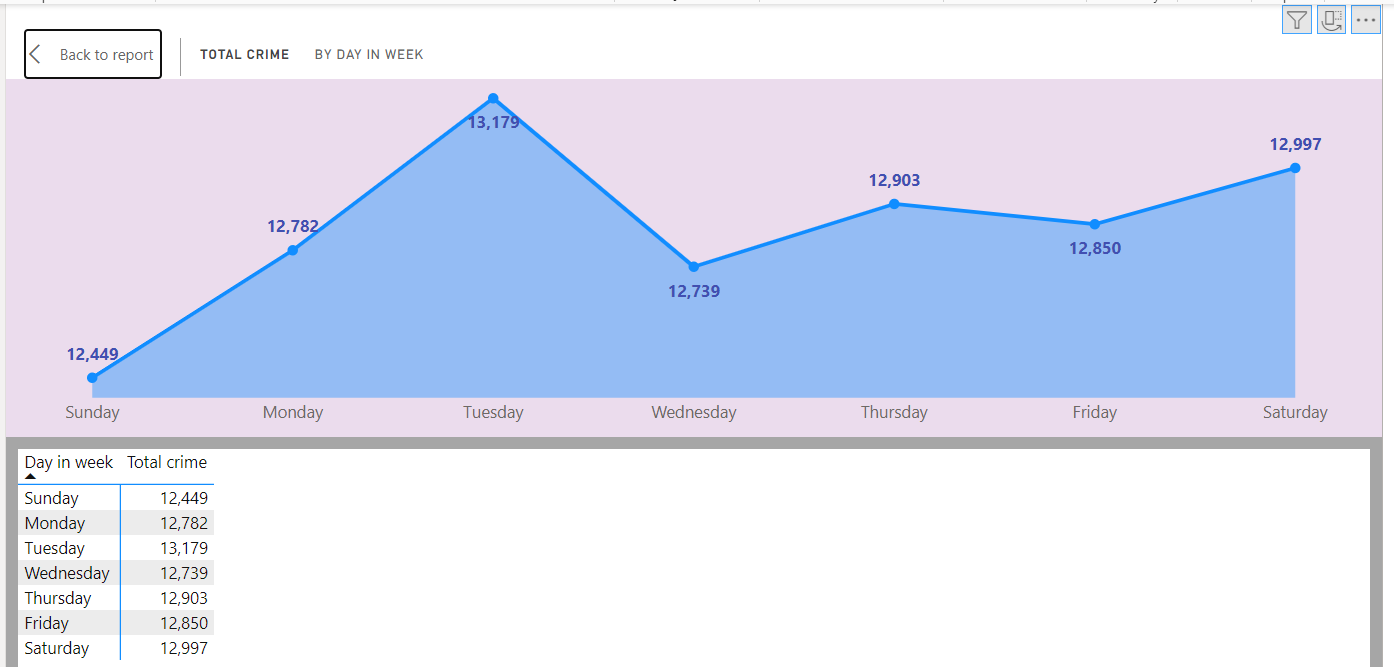
**Crime seems to be spread out across the city overall.** However, there's a big difference in how much crime happens during the day compared to night. Crime goes way up at night, so we need to focus on preventing it then.

**To stop crime at night, here are some ideas:**

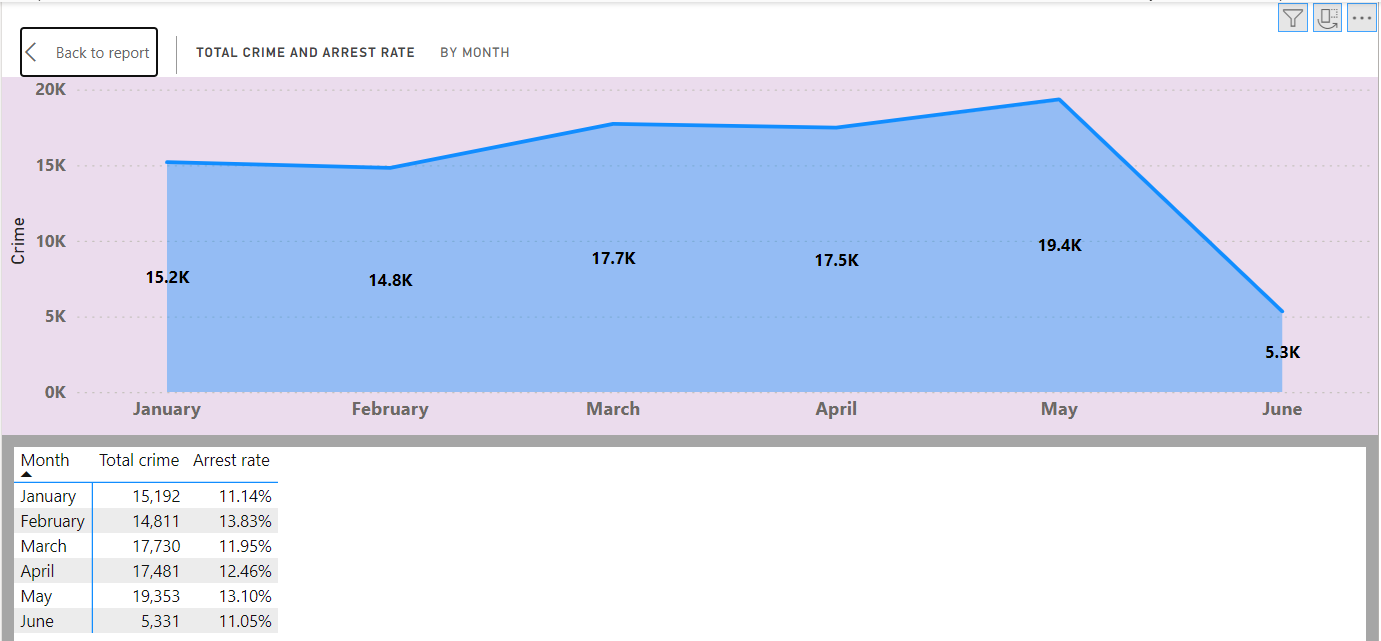
* More police patrols at night
* Better lighting in areas where people feel unsafe at night
* Programs to get residents involved in keeping their neighborhoods safe at night

By focusing on these night-time crime prevention strategies, the police can make the city safer overall.

* Weekly Trend



* Monthly Trend



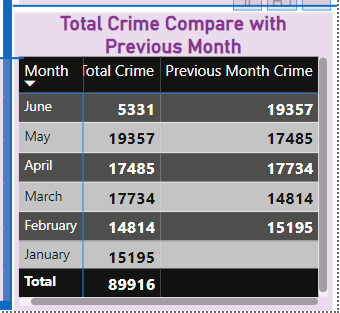
1. **Create a monthwise tabular data consisting of two columns, month and total no. of crimes in that month. Also, add one more column where each row of the column contains the total no. of crimes for the previous month. Do we need to use any filter-based DAX function here (All, All except, etc)?**

**Ans :-** **In my crime analysis project, I wanted to look at crime rates by month.** To do this effectively, I needed a special calendar in Power BI (a date table). But this calendar wouldn't let me easily compare crime rates between months (like total crimes in May vs. April).

So, I created a new table listing all the specific dates crimes happened. Then, I connected this new table to the calendar one-to-one. This trick allowed me to use a powerful function in Power BI (PREVIOUSMONTH) to calculate crimes in the previous month.

Finally, I used a table in Power BI to show both the current month's crimes and the previous month's crimes side-by-side. This way, I could easily see how crime rates changed from month to month.

By setting up my data this way, I could analyze how crime goes up and down throughout the year. This information is helpful for understanding crime patterns and making decisions about how to prevent crime.



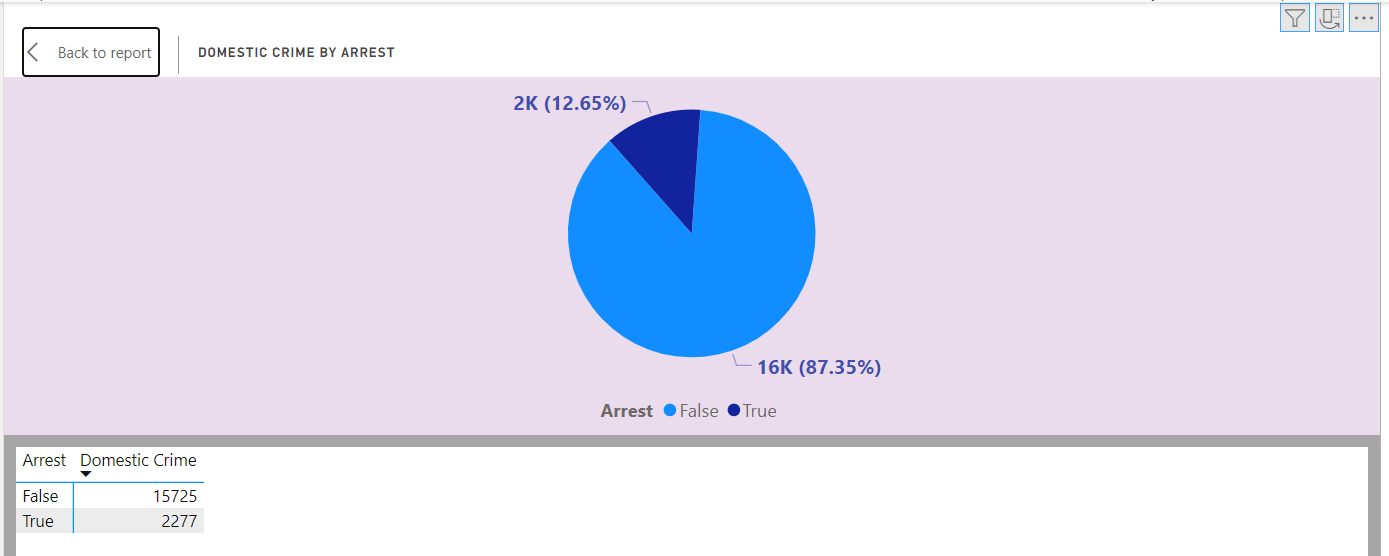
1. **As per the previous reports, most domestic crimes do not result in arrest due to public hesitation and family pressure, is this trend also visible in our data?**

**Ans :-** **The data shows a troubling problem with domestic violence arrests.** The pie chart reveals only 12.65% of domestic crimes lead to arrests. This isn't surprising - many reports say people hesitate to report domestic violence due to shame or family pressure.

**This low arrest rate is a major gap in how we fight domestic violence.** People might not report these crimes because they're afraid or pressured by family.

**Here's what we can do to fix this:**

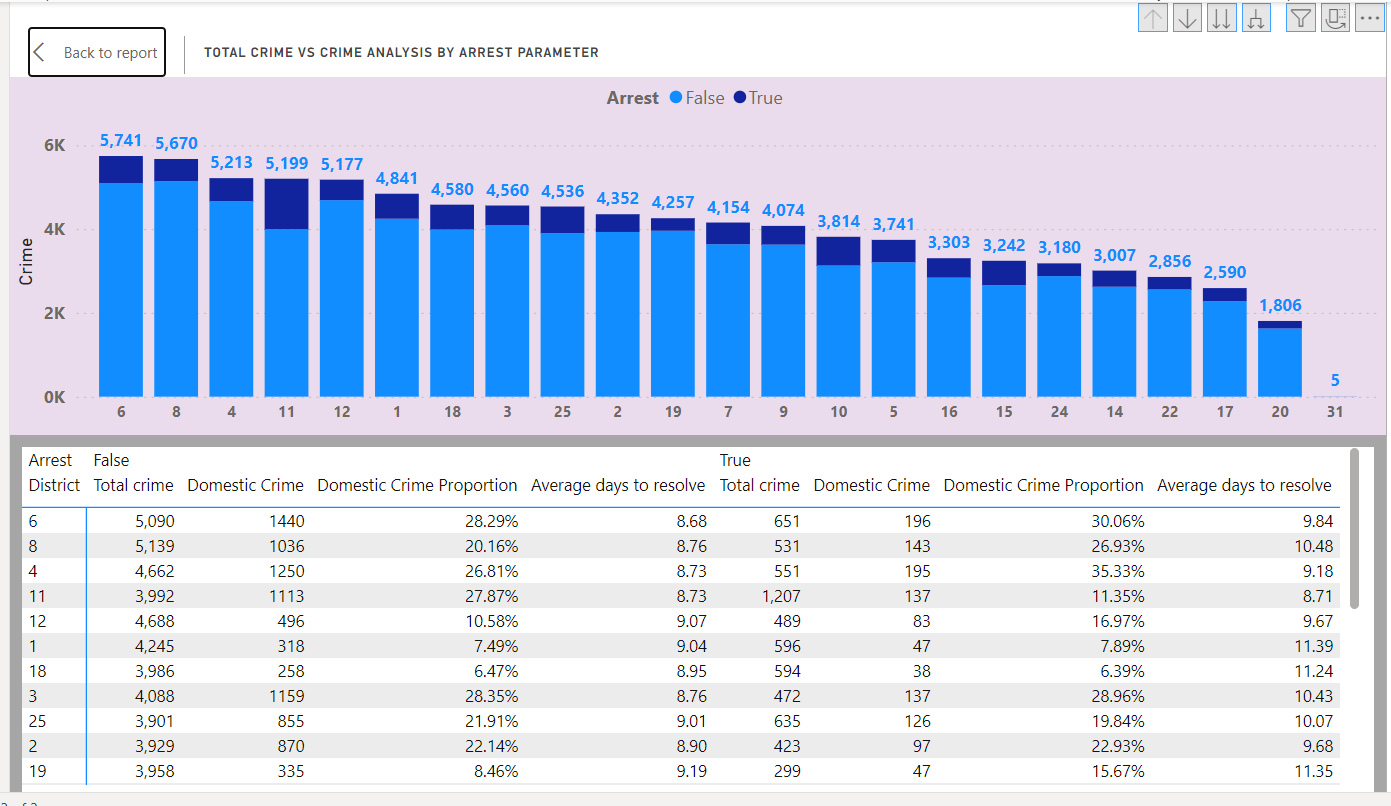
* **Encourage people to report domestic violence:** Make it easier and safer for people to report abuse.
* **Support victims:** Provide resources and help for people experiencing domestic violence.
* **Train law enforcement:** Train police officers to handle domestic violence cases with sensitivity and effectiveness.
* **Raise awareness:** Talk about the importance of reporting domestic violence and break the silence.

By taking these steps, we can increase arrest rates and hold abusers accountable. This will ultimately help create a safer environment for everyone.  


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

**Ans :-** **Taking a closer look at the descriptions of these crimes reveals some patterns.** The most common terms include 'Simple' (11,244 incidents), 'domestic Battery Simple' (8,060 incidents), '$500 And Under' (7,024 incidents), and 'over $500' (6,813 incidents). This helpful visualization lets us see what types of crimes are happening most often.

**Stakeholders can use this information to understand the bigger picture.** The visualization also allows you to filter by category for a more focused view, or remove the filter to see everything at once.



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

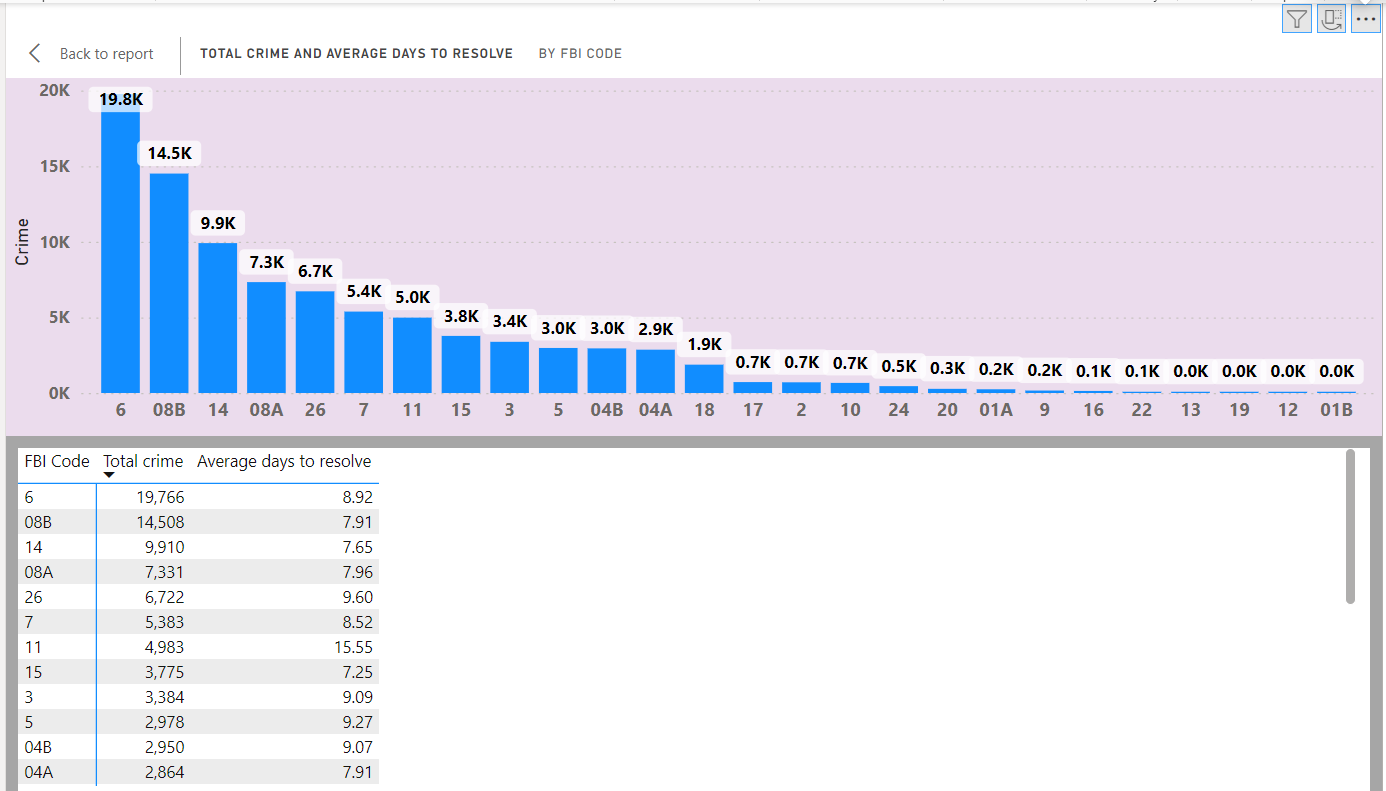
**Ans :-** **The data doesn't tell us exactly where domestic crimes happen.** Domestic violence isn't listed by location, only by the type of crime (like "domestic battery").

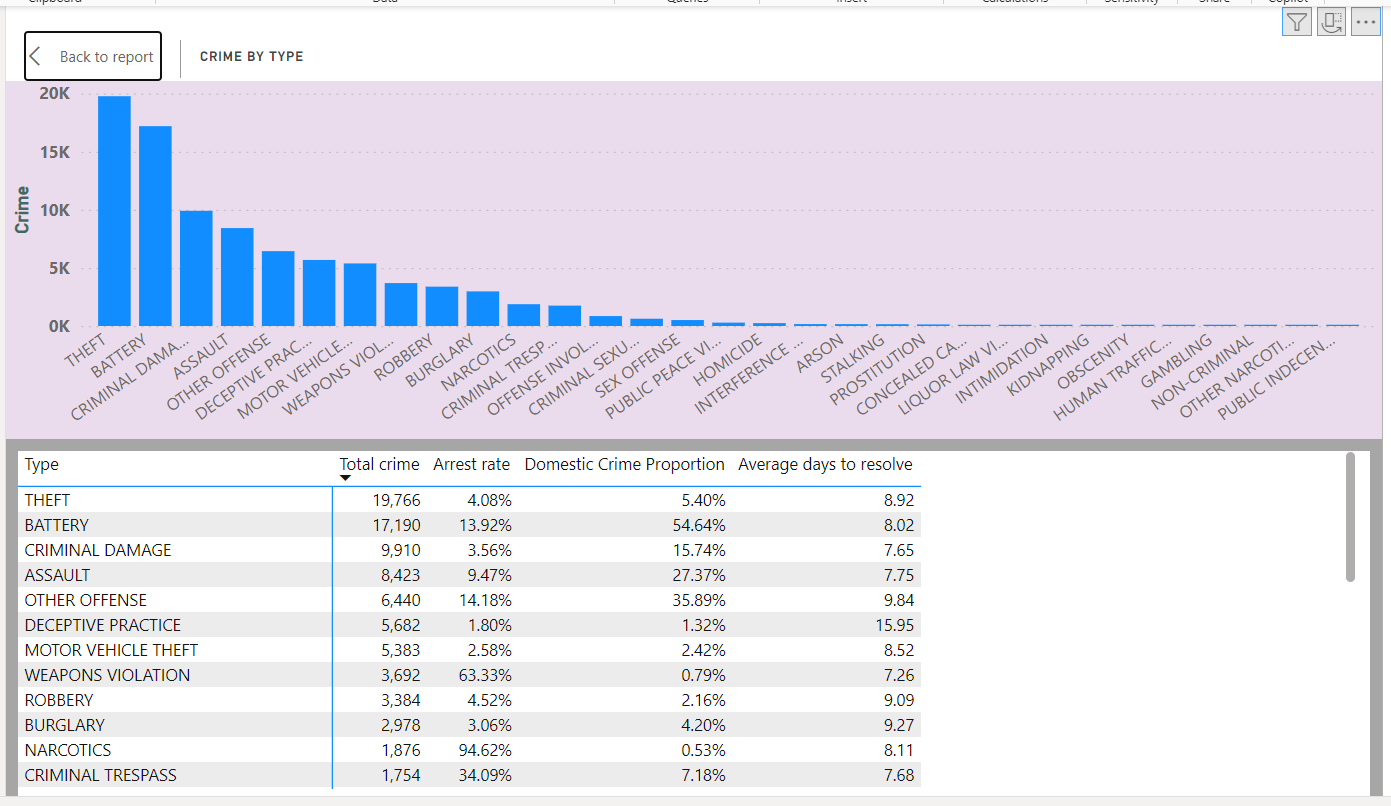
**To find areas with high rates of domestic violence, we'd need to look deeper.** We could analyze specific addresses or descriptions to see where these crimes are happening most often.

**Domestic violence can happen anywhere.** To tackle this issue, we need law enforcement, social services, and community groups to work together. They can help victims and hold offenders accountable.

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

**Ans :-** **This chart shows how quickly different crimes get resolved.** Crimes like traffic violations and gambling are resolved quickly, while trickier crimes like human trafficking take much longer. Similarly, cases at places like airports are resolved faster than cases at pawn shops or colleges.



**The chart also shows that some FBI codes (like 19, 15, and 14) are linked to faster case resolution.** This suggests that law enforcement has figured out good ways to handle these types of crimes. By understanding which crime codes tend to be resolved quickly, police forces can learn from these successes and improve their overall efficiency.  


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

**Ans :-** **I created a new column to help me find domestic violence crimes in District 8.** The formula checks two things: if the crime was domestic violence ("Domestic") and if it happened in District 8 ("District" = "8"). If both are true, the column says "Yes". If not, it says "No". This new column will be helpful for analyzing domestic violence specifically in District 8.

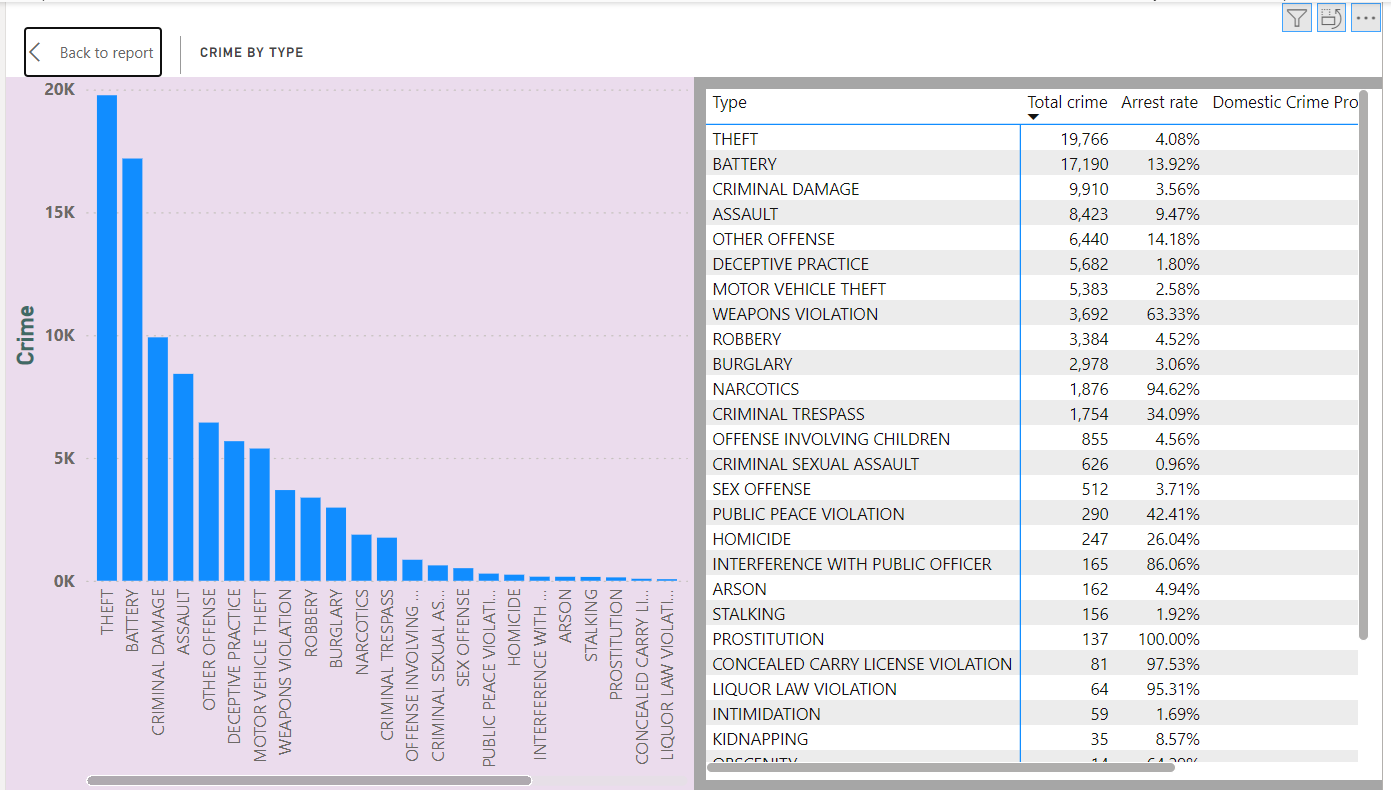
Domestic Crime in District 8 = IF( AND('crimes data 2022'[Domestic], 'crimes data 2022'[District] = "8"), "Yes", "No" )

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 :-** **Our data shows a lot of theft, battery, and vandalism.** These crimes are common, but the arrest rates are low.

**Theft is especially worrying.** It hurts people financially and makes our communities feel unsafe.

**Looking at how many of each crime happened, theft is probably the most common, followed by battery and vandalism.** This shows how serious these crimes are and how hard it can be to stop them.



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

**Ans :-** **Imagine Power BI as a cloud service for making data dashboards and reports.** But what if your data is stored on your own computers (on-premises) and not in the cloud? Power BI Gateway is like a bridge. It connects Power BI in the cloud to your on-premises data securely.

**Here's what Power BI Gateway can do:**

* **Connect to your data:** It lets Power BI talk to different data sources on your computers, like SQL Server or Oracle.
* **Keep reports up-to-date:** Power BI Gateway can refresh the data in your reports automatically, so you're always looking at the latest information.
* **Analyze the most recent data:** It can connect directly to your on-premises data without storing it in the cloud, so you can see the very latest details.
* **Two versions to choose from:** There's a Personal Gateway for individuals and small teams, and an Enterprise Gateway for bigger organizations with more users and complex data.

**Why would you use Power BI Gateway?**

* **Mix of cloud and on-premises data:** It helps if you have some data in the cloud and some on your own computers.
* **Always have fresh data:** Your reports can automatically refresh with the latest information.
* **See the latest data right away:** You can analyze data on your computers without waiting to copy it to the cloud first.
* **Keep data secure:** Sensitive data stays on your computers, which might be important for security reasons.
* **Large organizations:** The Enterprise Gateway helps manage Power BI for big companies with many users.

**In short, Power BI Gateway makes Power BI much more useful by letting it securely access and update your on-premises data.**

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

**Ans :-** **Here's my plan to analyze crime data:**

1. **Overall Crime Rates:** First, let's look at the total number of crimes and how many arrests were made. This will give us a general idea of how much crime there is and how good the police are at catching criminals.
2. **Crime by Location:** Next, let's use a map to see where crimes are happening. This will help us identify areas with high crime rates so we can focus our resources there.
3. **Day vs. Night Crime:** We should also see what time of day most crimes happen (day or night). This will help us figure out when to schedule police patrols.
4. **Crime Types in Different Areas:** Let's take a closer look at the different types of crimes happening in different parts of the city. This can help us understand crime patterns and target our prevention efforts.
5. **How Long Cases Take to Solve:** We can calculate how long it takes the police to solve crimes. This will show us how efficient they are and if there are areas for improvement.
6. **Domestic Violence:** Finally, let's look at domestic violence crimes in particular. We need to understand why these crimes are different and how to best support victims.

By following these steps, we can gain valuable insights from the crime data and use them to make our city safer.

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?**

**Ans :-** If we add a "district ID" to the new table, it creates a special connection with the existing table. Imagine each district is like a store, and the cases are like products. A single district (store) can have many cases (products), but each case (product) belongs to only one district (store). That's why it's called a many-to-one relationship.