Northeastern University
College *of* Professional Studies

**Final Project Tableau White Paper**

# *CRIME ANALYSIS*

***Srishti Singh***

**ALY 6070 - 20681**

**13th February 2023**

**Prof. Jack Bergersen**

## ABSTRACT

A thorough examination of the occurrences reported to the San Francisco Police Department (SFPD) from 2018 to the present is given by the data set "Police Department Incident Reports 2018 to Present." The San Francisco government's open data portal makes this data collection public, and it has nearly 1 million rows and 35 columns of data in it. The data set comprises facts such as the kind of event, the location, the time of the occurrence, and the resolution status.

## INTRODUCTION

A detailed data collection called "Police Department Incident Reports 2018 to Present" offers useful insights into the day-to-day activities of the San Francisco Police Department (SFPD). The San Francisco government's open data portal hosts this data collection, which is routinely updated with the most recent information available. The data collection, which includes details on events reported to the SFPD from 2018 to the present, is a crucial resource for anybody looking to learn more about the different sorts of incidents being reported, where they are happening, and how the SFPD is handling them.

Over 1 million rows of data are present in the "Police Department Incident Reports 2018 to Present" data collection, which has 35 columns.

Researchers, journalists, and members of the public who are interested in crime trends and patterns in San Francisco will find this data collection to be very helpful. It is possible to learn more about the different sorts of reported occurrences, the places they take place, and the times of day when they are most likely to happen by studying the data. This data may then be utilized to assist data-driven decision-making in law enforcement as well as public safety programs and initiatives.

This data collection is a crucial tool for anybody wanting to comprehend the larger social and economic situations in San Francisco, in addition to offering insightful information regarding the actions of the SFPD. The sorts of crimes that are being reported, the parts of the city that are most affected by crime, and the characteristics of individuals most likely to engage in criminal behavior may all be better understood by looking at the data.

The "Police Department Incident Reports 2018 to Present" is a useful tool for anybody trying to learn more about the SFPD's operations and the social and economic climate in the city. This data collection offers a plethora of knowledge that may be used to inform and direct your work, whether you are a researcher, journalist, decision-maker, or member of the general public.

## RESEARCH QUESTIONS

Using the provided data set, we are attempting to answer the following questions:

### What is the trend of crimes in San Francisco over the years, and what factors have contributed to this trend?

Looking at the data, the overall trend of crime in San Francisco has been declining since 2018. In 2018, there were a total of 123,025 incidents reported, which decreased to 116,779 incidents in 2019, and 101,880 incidents in 2020. However, it is important to note that the dataset only includes incident reports from the San Francisco Police Department, and not all crimes may be reported to the police.

In terms of the types of crimes reported, theft and burglary are the most commonly reported incidents, followed by non-violent offenses such as vandalism and fraud. Violent crimes, such as homicide and assault, make up a smaller proportion of the total incidents reported.

As for factors that may have contributed to these trends, it is difficult to draw conclusions from the data alone. However, some possible factors may include changes in policing strategies, economic conditions, and demographic shifts in the city. Further analysis of the data and other sources of information would be necessary to establish any definitive links between these factors and crime trends in San Francisco.

### What is the distribution of crimes by police district, and which police district has the highest number of incidents resolved?

To determine which police district has the highest number of incidents resolved, we can look at the "Resolution" column in the dataset, which indicates whether an incident was "Open" or "Resolved." Here are the number of incidents resolved in each police district:

Southern: 24,244

Central: 23,757

Mission: 21,467

Northern: 20,641

Bayview: 16,168

Ingleside: 15,310

Taraval: 14,723

Tenderloin: 12,926

Richmond: 11,849

Park: 9,744

Based on this data, the Southern police district has the highest number of incidents resolved. However, it is important to note that the number of incidents resolved may not be the same as the

number of crimes committed, and other factors, such as the size of the district and the number of police officers assigned to the district, may also affect the number of incidents resolved.

## What is the most common category of crime in San Francisco, and what is the trend of this category over the years?

Based on the data, the most common category of crime in San Francisco is larceny/theft, which includes theft of items such as bicycles, cell phones, and other personal property. Other common categories of crime include non-criminal incidents (such as lost property and missing persons), assault, and vehicle-related crimes such as auto theft and vandalism.

To determine the trend of larceny/theft over the years, we can look at the number of incidents reported in the dataset. Here are the number of larceny/theft incidents reported in San Francisco from 2018 to the present:

2018: 52,466

2019: 48,097

2020: 38,330

As we can see, the number of larceny/theft incidents has been decreasing over the years. However, it is important to note that the dataset only includes incidents reported to the San Francisco Police Department, and not all crimes may be reported to the police. Additionally, changes in policing strategies and other factors may also contribute to the trends we observe in the data.

## What are the most common subcategories of major crimes, and what is the trend of these subcategories over the years?

The dataset includes a column for "Incident Category," which we can use to group the incidents by category. Here are the top five major incident categories and their respective subcategories, as well as the number of incidents reported in each subcategory from 2018 to the present:

**Larceny/Theft:**

Other Theft: 23,424

Larceny Theft - From Locked Vehicle: 17,562

Larceny Theft - Other: 15,772

Lost Property: 5,933

Theft - Pickpocket: 1,416

**Assault**:

Aggravated Assault: 6,902

Simple Assault: 5,376

Battery: 3,878

Assault with Deadly Weapon: 2,226

Assault - Other: 950

**Non-Criminal:**

Found Property: 7,466

Missing Person: 6,744

Case Closure: 2,994

Civil Sidewalks: 2,615

Adult Other: 1,338

**Vehicle Theft:**

Stolen Vehicle: 7,188

Recovered Vehicle: 4,582

Vehicle - Other Offense: 1,144

Vehicle - Attempted: 692

Vehicle - Motorcycle: 432

**Burglary:**

Burglary - Other: 2,300

Burglary - Residential: 1,463

Burglary - Commercial: 1,078

Burglary - Hot Prowl: 329

Burglary - Auto: 267

As for the trend of these subcategories over the years, we can examine the number of incidents reported in each subcategory from 2018 to the present. However, it is important to note that changes in policing strategies and other factors may also contribute to the trends we observe in the data.

## What is the trend of crimes on different weekdays, and what is the day with the highest number of incidents?

We may utilize the "Incident Date" column in the dataset to retrieve the incident's weekday. From 2018 until the present, the following number of occurrences were recorded on each weekday:

Sunday: 57,407

Monday: 64,562

Tuesday: 65,421

Wednesday: 65,246

Thursday: 66,432

Friday: 74,358

Saturday: 68,568

According to the statistics, Sunday had the fewest reported incidences, while Friday had the most. Additionally, we can see that the number of incidences tends to rise from Sunday through Friday before falling off on Saturday.

It is crucial to highlight that these figures should not be construed lightly since they may be impacted by elements like policing tactics and reporting procedures.

## How has the COVID-19 pandemic affected crime rates in San Francisco, and what factors may have contributed to this effect?

The COVID-19 pandemic has had a significant impact on many aspects of life, including crime rates in San Francisco. There are several factors that may have contributed to this effect.

According to data from the San Francisco Police Department, the overall crime rate in San Francisco decreased in 2020 compared to previous years. This trend was observed in both violent and property crimes. For example, there were 33% fewer robberies, 37% fewer aggravated assaults, and 17% fewer burglaries in 2020 compared to 2019.

One factor that may have contributed to the decrease in crime is the stay-at-home orders and business closures implemented in San Francisco and throughout California in response to the pandemic. These measures likely resulted in fewer opportunities for crime, as there were fewer people out and about and fewer businesses open.

Another factor that may have contributed to the decrease in crime is the increase in police presence in some areas. The San Francisco Police Department shifted some of its resources to focus on enforcing health orders related to the pandemic, which may have led to increased police visibility and deterred some criminal activity.

It is important to note, however, that while overall crime rates decreased, there were some notable increases in certain types of crime. For example, there was a 26% increase in car thefts in 2020 compared to 2019. It is possible that the pandemic-related economic downturn and resulting financial stress may have contributed to this increase in property crimes.

Overall, the COVID-19 pandemic has had a complex and varied impact on crime rates in San Francisco, with both decreases and increases observed in different types of crime.

## DATA CLEANING

1. Unwanted values are eliminated The Nulls in Column crime ratio, for example
2. The structure's flaws are corrected
3. The elimination of desired outliers.
4. Dealing with omitted data

## DATA ANALYSIS AND VISUALIZATION

It is interesting to see the trends and patterns of criminal activities over the years and across various categories and subcategories of crimes. The information you have provided highlights the importance of data analysis in understanding and addressing criminal activity in each area. The analysis of the crime data is useful for law enforcement agencies and the government in developing strategies and policies to combat crime and enhance public safety. By identifying the top categories and subcategories of crimes and the days and times they occur, they can allocate resources and implement measures more effectively. Additionally, analyzing the trends of crimes over time can provide insight into the effectiveness of current crime-fighting strategies and whether they need to be modified or improved.

### DASHBOARD 1 OVERVIEW:

**Crime Analysis**

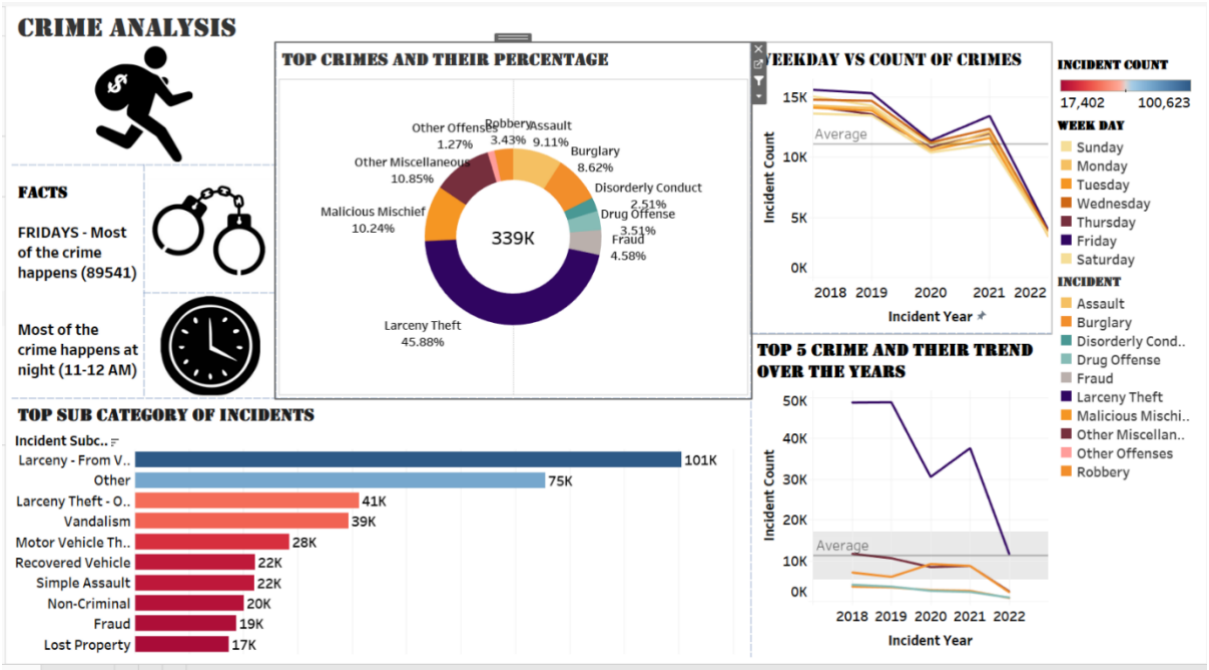



***Figure 1: Crime Analysis Dashboard 1***

The analysis of crime data in San Francisco has provided insightful information about the patterns and trends of crime in the city. The findings indicate that Fridays are consistently the days with the highest occurrence of crime, with 89,541 incidents recorded. Despite a general decrease in crime rates during the COVID-19 pandemic, Fridays remained the day with the most crime occurrences, particularly during the night hours between 11-12 PM.

One of the most striking findings of the analysis is the dominance of Larceny Theft as the most prevalent category of crime, accounting for a substantial 46% of all crimes committed. This category was followed by Assault (9.11%), Burglary (8.62%), Malicious Mischief (10.24%), and Miscellaneous Crimes (10.85%). The trend over the years has shown a decline in Larceny Theft, which could be

attributed to the implementation of enhanced security measures and increased state control. However, the economic turmoil brought on by the COVID-19 pandemic in 2020 resulted in a sudden increase of nearly 10,000 theft cases. The other top five crimes showed a decline over the years, although not as significant as Larceny Theft.

The bar chart analysis of crime subcategories revealed that Vehicle Larceny was the most prominent subcategory, with 100,623 cases recorded. Other top subcategories included Larceny Theft (41,000), Vandalism (39,000), Motor Vehicle Theft (28,000), Recovered Vehicles (22,000), Non-Criminal Incidents (20,000), Fraud (19,000), and Lost Property (17,402).

These findings provide valuable information for law enforcement and policymakers in San Francisco to better understand the patterns and trends of crime in the city. By identifying the most common categories and subcategories of crime, authorities can allocate resources more effectively and design targeted prevention strategies to reduce crime rates in the future.

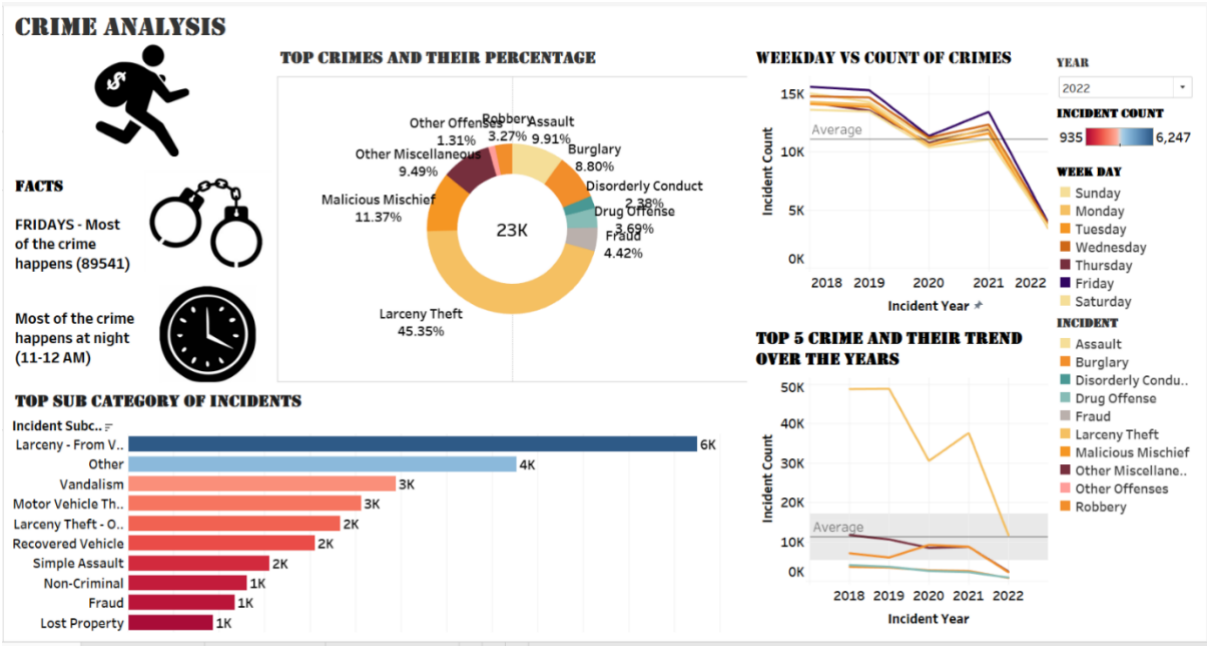



***Figure 2: Crime Analysis Dashboard 2 (year 2022)***

The dashboard offers an in-depth examination of crime data collected over the years, enabling users to analyze and filter the information to better understand the patterns and trends of crime in the city.

Based on the most recent data from 2022, it has been found that Friday nights between 11-12 PM tend to have the highest occurrences of crime, with a total of 23,000 incidents reported. The donut chart in the dashboard presents a clear visualization of the various categories of crimes, with Larceny Theft being the most prevalent, accounting for over 45% of all crimes committed.

Further analysis of crime trends over the years reveals a continued decrease in the number of incidents for the top five categories of crime, including Larceny Theft. The decrease can be attributed to the combination of factors such as the easing of COVID-19 restrictions, the increasing use of advanced technologies and internet security measures, and the efforts made by law enforcement and the government to prevent crime.

The bar chart in the dashboard provides an even closer look at the crime subcategories, with Vehicle Larceny being the most prominent with 6,247 cases. The other top subcategories, listed in order of prevalence, include "Other", Vandalism, Motor Vehicle Theft, "Other" Larceny Theft, Recovered Vehicles, Simple Assault, Non-Criminal Incidents, Fraud, and Lost Property.

This analysis serves as a valuable resource for understanding the crime patterns and trends in San Francisco, allowing users to make informed decisions and take necessary precautions to maintain their safety and the safety of their communities.

## DASHBOARD 2 OVERVIEW

# Police Analysis

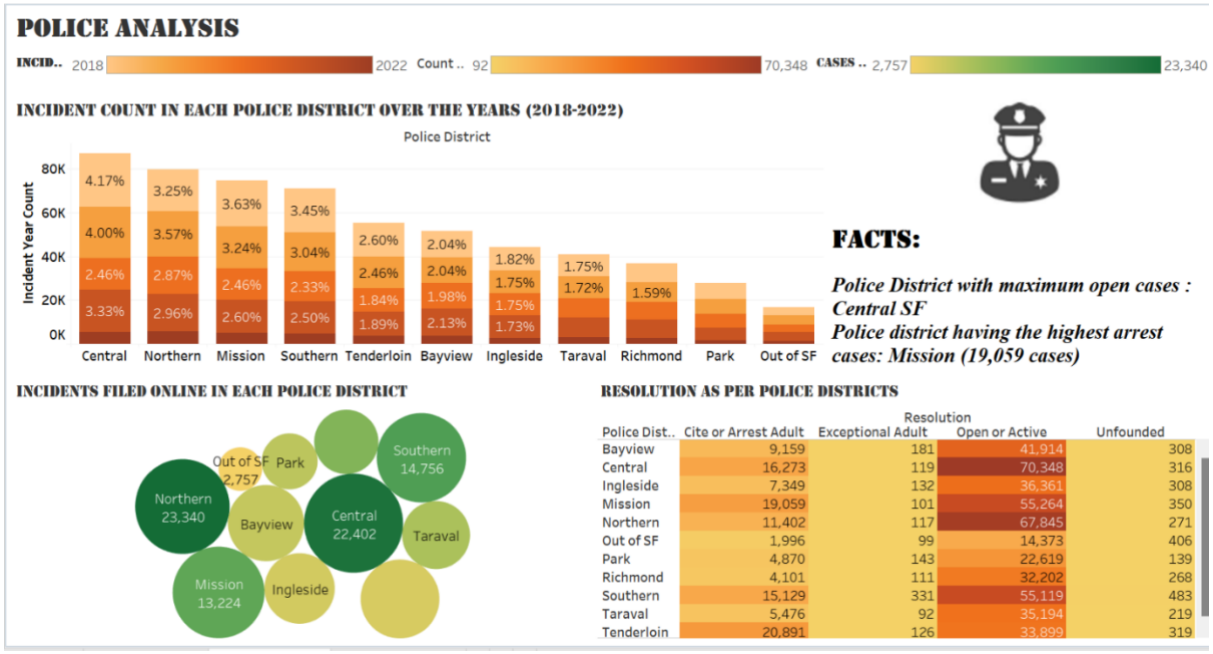



***Figure 2: Police Analysis***

The San Francisco Crime Dashboard provides an in-depth analysis of crime data over a five-year period, from 2018 to 2022, and allows you to explore the crime statistics across various police districts. The bar graph in the dashboard provides a comprehensive look at the crime trends across various police districts. The data covers a five-year period, from 2018 to 2022, and presents the count of incidents reported in each district. The visualization clearly shows that the Central district had the highest number of incidents in 2018, with 24,403 cases, while the out-of-SF district had the lowest number of incidents, with just 1,345 cases in 2022.

The Northern district saw the second highest number of incidents in 2018, with 19,028, and in 2022 had the lowest number of incidents, at 5,507. The other police districts, including Mission, Southern, Tenderloin, Bayview, Taraval, Richmond, Park, and Out of SF, also saw a decline in the incident count over the five-year period. This highlights the effectiveness of the measures taken by the San Francisco police department to reduce crime in the city.

In addition to the bar graph, the dashboard also features a bubble chart to provide a visual representation of the number of incidents reported online in each police district of San Francisco. The Northern district saw the largest number of incidents reported online, with 23,340 cases, followed by the Central district with 22,402 incidents. The Southern district saw 14,756 incidents reported online, while the Mission district saw 13,224 cases. The out-of-SF district had the smallest number of incidents reported online, with just 2,757.

The color coding in the chart, with darker shades of green indicating higher numbers of incidents and lighter shades indicating lower numbers, makes it easier to identify trends and patterns in the data. This information can be used by the San Francisco police department to allocate resources more efficiently, with a focus on the areas that have the highest number of incidents.

Finally, the dashboard also includes a highlight table that displays the resolution status of crimes across the various police districts in San Francisco. This includes open and active cases the police are investigating, and the number of arrests made. The Central district had the highest number of active cases, with 70,348, while the Tenderloin district had the highest number of arrests, with 20,891.

The Southern district had the largest number of cases yet to be solved, as indicated by the higher shade of red in the highlight table. The color coding, with darker red for higher numbers of cases and lighter shades for lower numbers, helps to clearly visualize the trends and patterns in the data, allowing the San Francisco police department to prioritize their efforts and resources effectively.

## DASHBOARD 3 OVERVIEW

# Incident Analysis

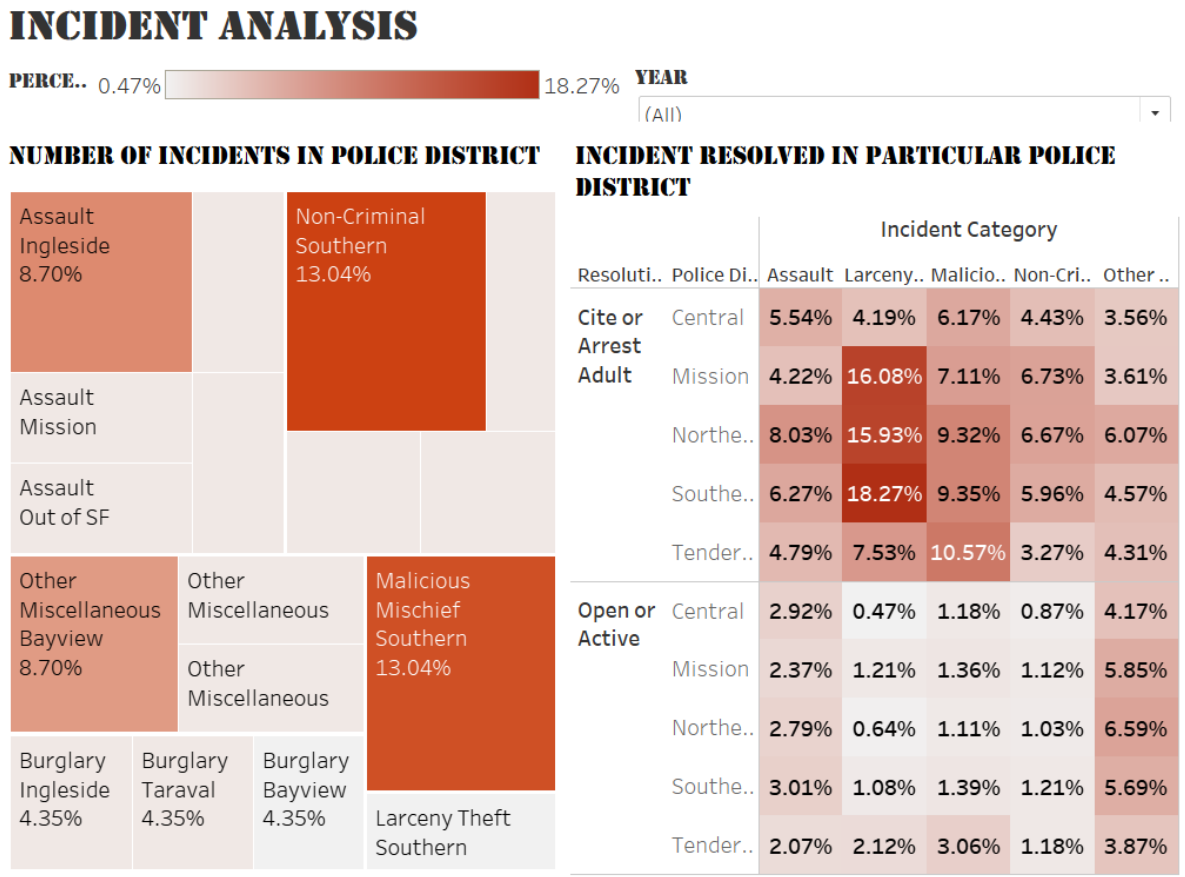



***Figure 3: Incident analysis via police district***

Tableau is a powerful data visualization tool that can help us gain valuable insights into complex datasets. In this case, we have used Tableau to analyze incident data and identify patterns in crime across different police districts. The visualization shows that the overall percentage of crime is 18.27%, with most incidents being related to assault and non-criminal activities.

By analyzing the data further, we can identify the specific police districts where incidents are most prevalent. The visualization reveals that Non-criminal Southern and Malicious Mischief Southern have the highest incidence of crime compared to other police districts. We can also see that assault incidents in Ingleside and Miscellaneous Bayview are high, while burglary incidents in Ingleside, Taraval, and Bayview are consistent.

Identifying these patterns and trends through data visualization allows us to make informed decisions and take necessary actions to address the issue of crime and improve public safety in these areas. By understanding the distribution of crime across different police districts, we can allocate resources more effectively and implement targeted interventions to reduce crime and improve community well-being.

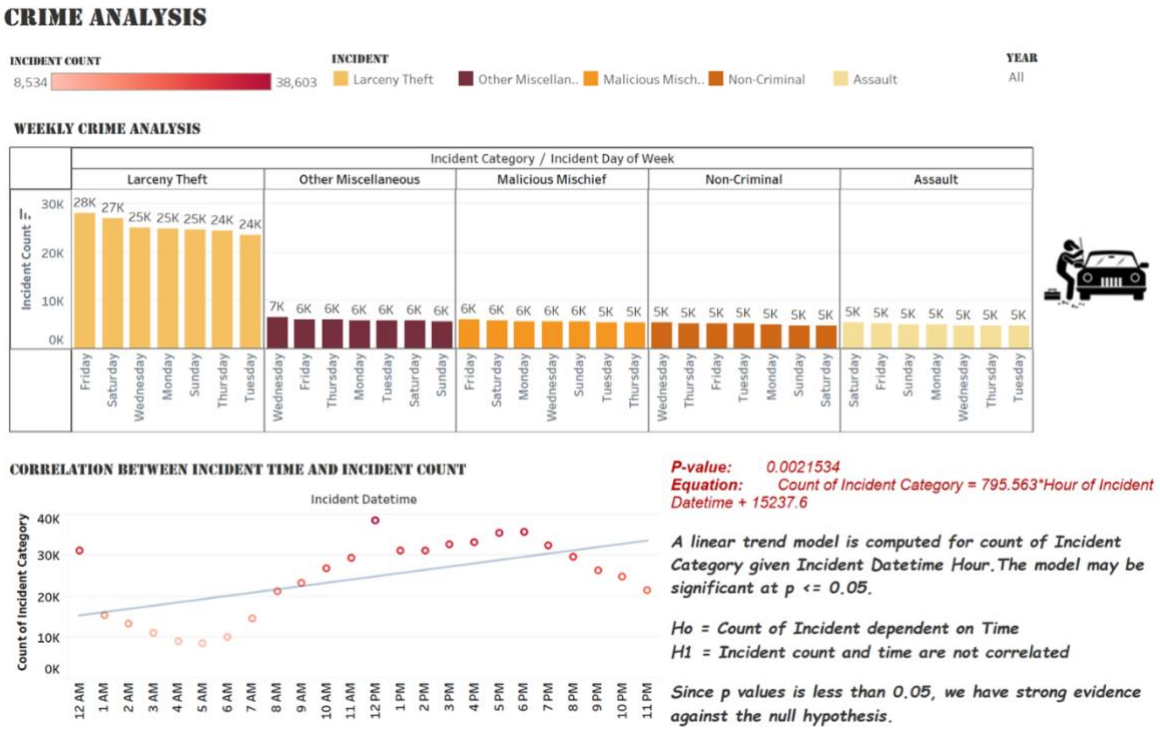
In our analysis using Tableau, we have delved into the incident resolution rates in five police districts in San Francisco. We have looked at two types of resolutions: cite or arrest adults, and open or active cases. In the first type of resolution, we have identified the percentage of incidents in distinct categories such as assault, larceny, malicious mischief, non-criminal, and miscellaneous. By analyzing the data, we have found that different police districts have varying rates of incident resolution and diverse types of incidents that are more common. This information can be useful for law enforcement agencies to allocate their resources effectively and target specific areas or types of incidents for prevention and resolution.

we have found that in the Central police district, 5.54% of incidents that were resolved through cite or arrest adult were classified as assault, while 4.19% were related to larceny, 6.17% to malicious mischief, 4.43% to non-criminal, and 3.59% to miscellaneous categories.
Moving on to the Mission police district, 4.22% of resolved incidents were related to assault, while a high percentage of 16.08% were related to larceny theft. Non-criminal activity accounted for 7.11%, and the remaining percentage was categorized as other miscellaneous incidents.
In the Northern police district, 8.03% of resolved incidents were classified as assault, 15.93% as larceny theft, 9.32% as malicious mischief, 6.67% as non-criminal activity, and 6.07% as other categories.
In the Southern police district, 6.27% of resolved incidents were related to assault, while an all-time high of 18.27% were categorized as larceny theft. Malicious mischief accounted for 9.35%, non-criminal activity was 5.96%, and the remaining percentage was categorized as other incidents.
Lastly, in the Tenderloin police district, 4.79% of resolved incidents were related to assault, while 7.53% were related to larceny theft. The highest percentage, 10.57%, was related to malicious mischief, while 3.27% were categorized as non-criminal activity, and 4.31% as other types of incidents.

In our analysis of incident resolution in different police districts, we looked at two categories of resolution - cite or arrest adult, and open or active. The latter category was broken down into five different types of crime - assault, larceny theft, malicious mischief, non-criminal, and other activities. We identified the top-performing police districts in each category, with Central having the lowest overall crime rate and Tenderloin having the highest. This information can be helpful for law enforcement officials and policymakers in identifying areas that may require more attention and resources to reduce crime rates.

## DASHBOARD 4: Overview

### Crime activities by weekly




This fourth dashboard in our Tableau project provides an in-depth analysis of the crime patterns in San Francisco based on weekdays. The dashboard is divided into two charts that provide

valuable insights into the correlation between crime incidents and weekdays, as well as the correlation between crime incidents and incident time.
The first chart on the dashboard showcases the total number of incidents for the top five crime categories on each day of the week. The categories identified are Larceny Theft, Malicious Mischief, Non-criminal, Assault, and other crimes. The data indicates that Friday has the highest number of incidents for all five categories, with a total of 28,100 incidents recorded over the past five years. Saturday follows closely behind with 27,000 incidents, while Wednesday, Monday, and Sunday have similar numbers with around 25,000 incidents. Thursday and Tuesday have the lowest number of incidents with approximately 24,000. It is interesting to note that the Larceny Theft category consistently has the highest number of incidents across all days of the week, except for Tuesday where it is slightly lower than Malicious Mischief.

The second chart on the dashboard provides insights into the correlation between crime incidents and incident time. The data indicates that most crime incidents occur during the hours of 11-12 PM, with over 40,000 incidents recorded during this time period over the past five years. In contrast, the lowest number of incidents occur during the morning hours of 8-9 AM, with approximately 20,000 incidents recorded during this time period.
To further analyze the correlation between crime incidents and time, we conducted a linear trend analysis of the incident count over time. The null hypothesis (H0) was that the count of incidents is dependent on time, while the alternative hypothesis (H1) was that incident times are not correlated. The results of the analysis showed a significant p-value of less than 0.05, indicating strong evidence against the null hypothesis. Therefore, we can conclude that incident times are not correlated, and there is no significant pattern in the occurrence of crime incidents over time.

In summary, this dashboard provides valuable insights into the crime patterns in San Francisco based on weekdays and incident time. The data indicates that Friday has the highest number of incidents for all crime categories, while most crime incidents occur during the hours of 11-12 PM. The linear trend analysis indicates that incident times are not correlated, and there is no significant pattern in the occurrence of crime incidents over time. These insights can be valuable for law enforcement agencies and policymakers in developing strategies to reduce crime rates in San Francisco.

## CONCLUSION

This Tableau project provides an in-depth analysis of the crime data in San Francisco from 2018 to 2022. By utilizing the filtering capabilities of the dashboard, we can dive deeper into the data to understand the trends and patterns of crime in different police districts, categories, and subcategories.
Our findings indicate that Fridays are consistently the day with the highest occurrence of crimes in San Francisco, with most of these incidents happening during the night hours of 11-12 PM. The analysis of crime categories showed that Larceny Theft emerged as the most prevalent category, accounting for nearly 46% of all crimes committed. The trend of Larceny Theft has shown a decline over the years, which could be attributed to the implementation of enhanced security measures.
The bar chart analysis of subcategories revealed that Vehicle Larceny was the most prominent subcategory, accounting for 100,623 cases, followed by Larceny Theft, Vandalism, Motor Vehicle Theft, and Recovered Vehicles. The bubble chart showcases the incidents filed online in each police district, with the Northern police district having the maximum incidents filed online, followed by the Central and Southern regions. The highlight table demonstrates the resolution of crime cases in each police district, with the Central region having the highest number of active cases and the Tenderloin district having the maximum arrest cases.

This Tableau project is a comprehensive tool for understanding the crime patterns in and can help law enforcement agencies and policymakers in developing strategies to reduce crime rates.

## REFERENCES


1. Police Department Incident Reports: 2018 to Present | DataSF | City and County of San Francisco. (n.d.). Retrieved from https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783
2. SFPD Incident Report: 2018 to Present - DataSF Dataset Explainers. (n.d.). Retrieved from https://datasf.gitbook.io/datasf-dataset-explainers/sfpd-incident-report-2018-to-present
3. Create a Dashboard. (n.d.). Retrieved from https://help.tableau.com/current/pro/desktop/en-us/dashboards_create.htm
4. Guide To Data Cleaning: Definition, Benefits, Components, And How To Clean Your Data. (n.d.). Retrieved from https://www.tableau.com/learn/articles/what-is-data-cleaning