Crime Chronicles: Exploring Patterns and Trends in San Francisco

Introduction

San Francisco serves as a prominent commercial and financial hub within California, and it stands as the second most densely populated major city in the United States, following New York City. Notably, the city has encountered a concerning rise in its homeless population, with an increase of nearly 7,000 individuals since 2017, according to the Bay Area Economic Institute. This surge in homelessness has brought forth significant challenges, as a substantial portion of the city's homeless population grapples with mental illness and addiction, contributing to an escalation in criminal activity.

Furthermore, the advent of the pandemic has exacerbated issues of poverty and unemployment, which, in turn, can act as contributing factors to the prevalence of crime within the city. Consequently, our objective entails a comprehensive analysis of crime patterns in San Francisco through the utilization of Exploratory Data Analysis (EDA) and Time Series Analysis. By delving into these analytical approaches, we aim to discern significant patterns and trends within crime incidents, thereby enabling the police department to implement targeted measures in order to address the escalating crime rate effectively.

Data Description

The data used for this project has been gathered from an open data source on the internet : SFO Crime Data

It consists of 620K rows and 27 columns, where each row corresponds to a crime incident report filed by the police department. The crime reports are well documented and provide vital information such as Incident ID, Incident Number, Incident Date and Time, Crime Category and Subcategory, Place of incident and Resolution.

Data Cleaning

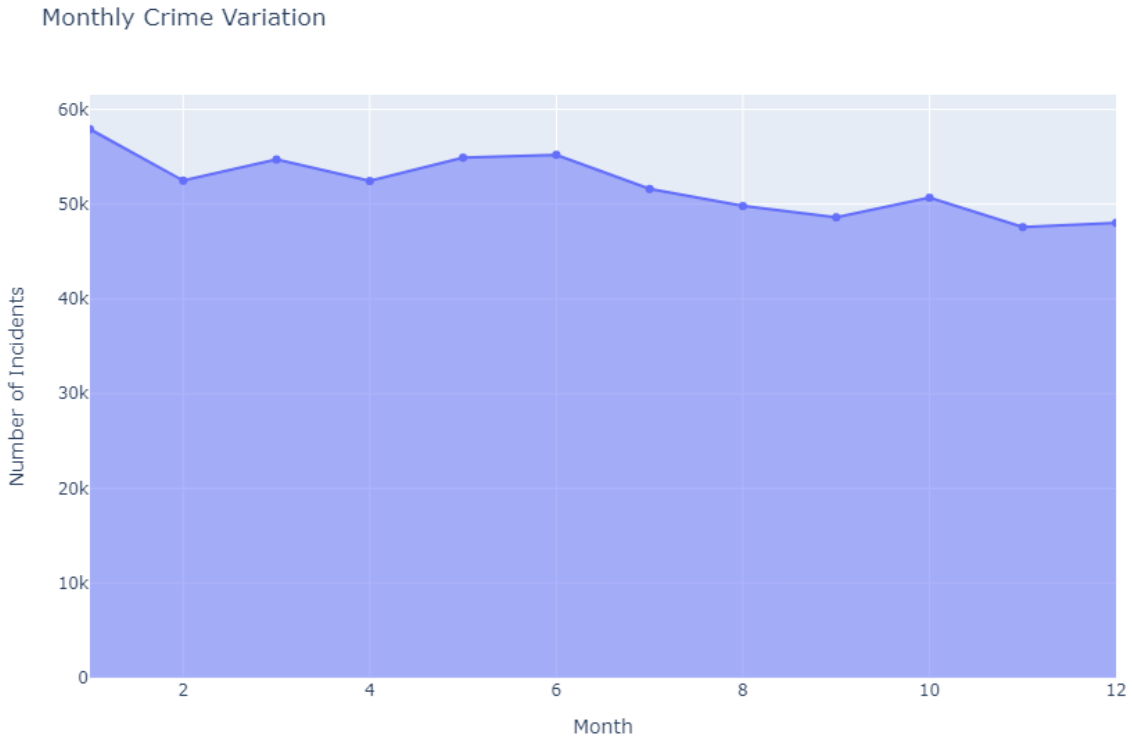
Out of the available 27 columns, we only use the below 13 columns for our analysis - IncidentID , IncidentNumber , IncidentMonth , IncidentYear , IncidentTime , IncidentDOW , IncidentCategory , IncidentSubcategory , Resolution, PoliceDistrict , Latitude , Longitude , ReportMethod. Below are the important steps performed as part of cleaning the data -

- We replace the NULL values under the ReportMethod as ‘Offline’.
- We update the ‘TRUE’ values under the ReportMethod as ‘Online’.
- NULL values under IncidentCategory and IncidentSubcategory are updated as ‘Unidentified’.
- The below values under IncidentCategory are merged into one:
    - Weapons Offense and Weapons Offence → Weapons Offense
    - Motor Vehicle Theft and Motor Vehicle Theft? → Motor Vehicle Theft
    - Other Miscellaneous and Other Offenses → Other
    - Suspicious Occ, Suspicious → Suspicious Occ

- The NULL values under Latitude and Longitude are replaced with the mean of Latitude and Longitude of corresponding PoliceDistrict values so that we confine the crime location to the district whilst holding on to the details that do not have Latitude and Longitude values.

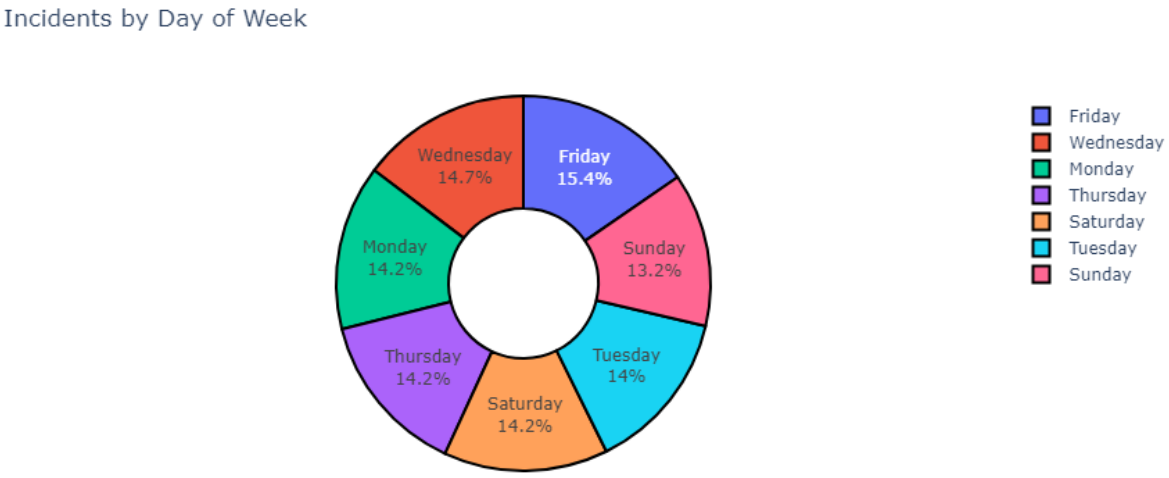
Key Insights and Findings

Insight 1: Monthly Crime Variation



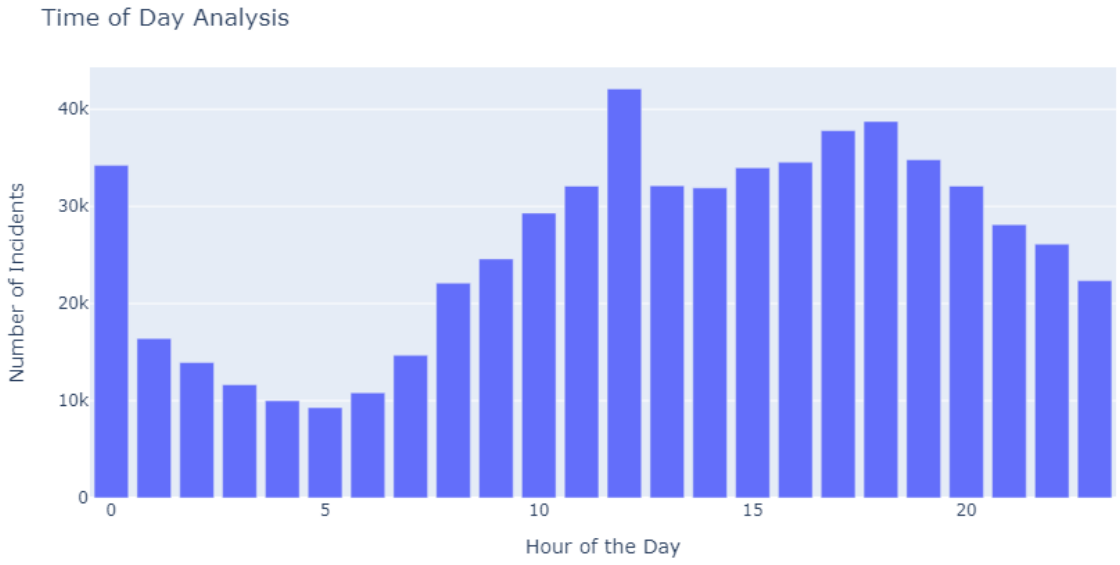
The above Area chart indicates that the highest number of crime incidents have taken place in January, followed by June and May. We also observe that the spread of incidents is almost equally distributed and there is no substantial peak or dip in the number of crimes for a specific month.

Insight 2: Day of the Week Crime Variation



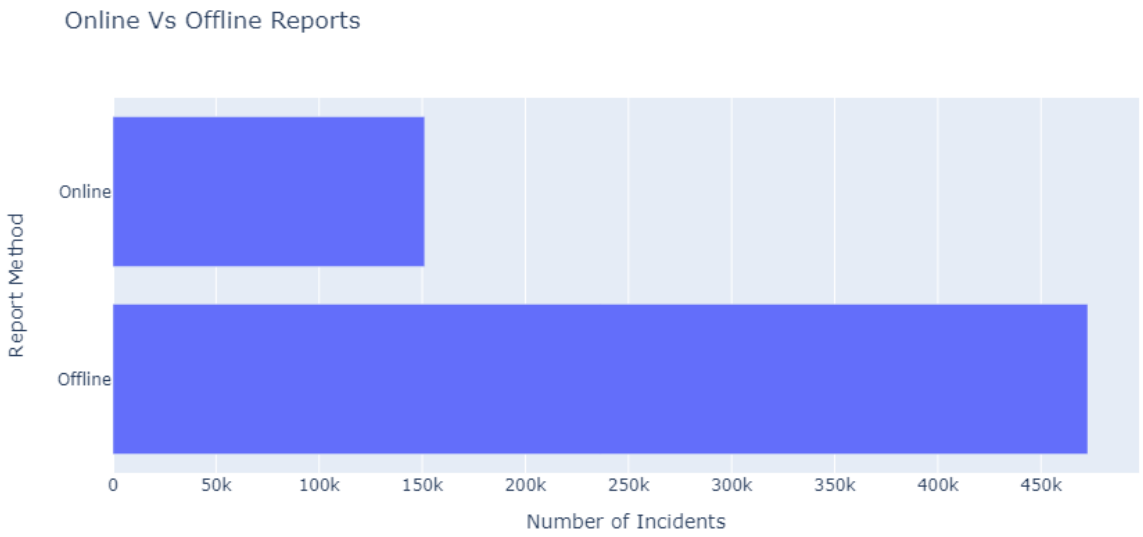
The above Pie Chart indicates that the highest crime incidents have been reported on Friday. Even though the crime rate seems to be uniform throughout the week, Fridays seem to be a little more susceptible to criminal incidents.

Insight 3: Time of the Day Analysis



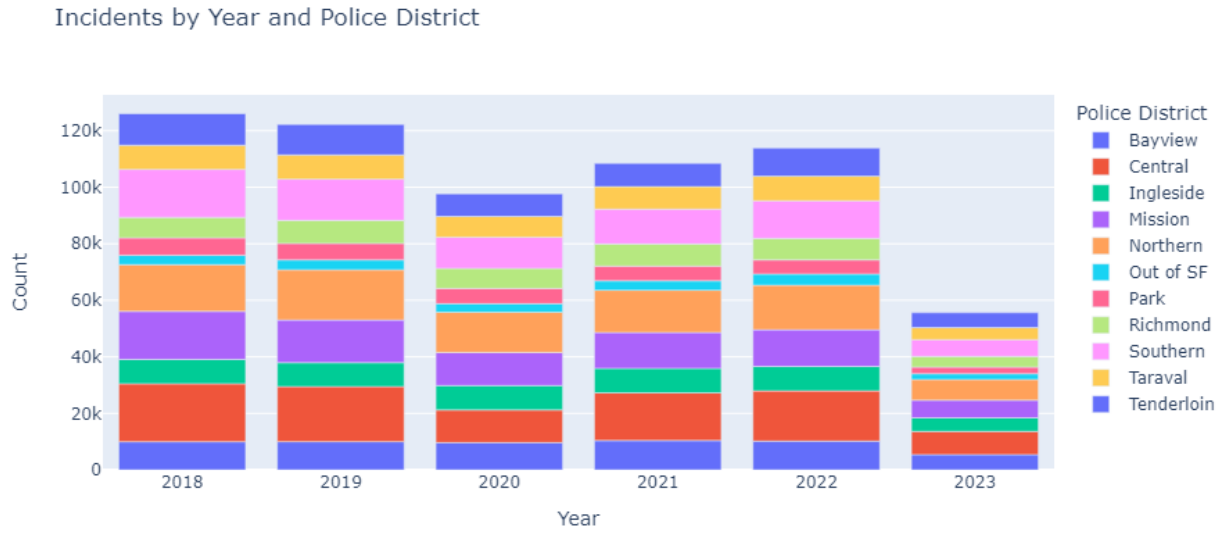
From the above Bar Chart, we can clearly notice that the crime incidents are significantly higher during the daytime, especially from mid-day to 6PM, when most people are at work.

Insight 4: Reporting Methods – Online vs Offline



From the above Bar Chart, it is quite evident that most of the crime incidents in San Francisco are being reported offline.

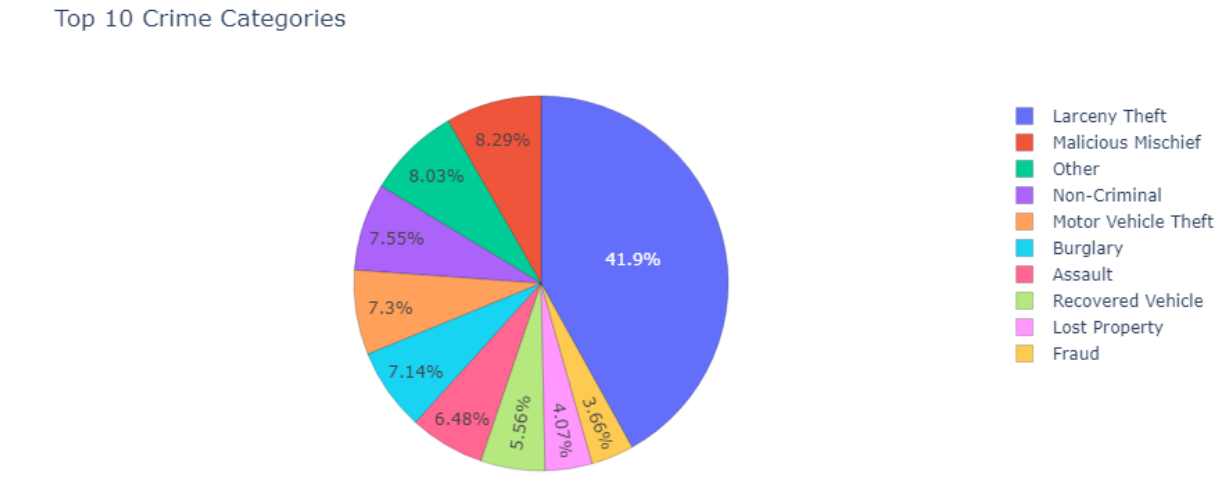
Insight 5: Year-wise Crime Counts per District



Above is the stacked bar chart that groups the number of crime incidents per district and arranges them year-wise to help us understand which district is more vulnerable to crime over the years. We can observe that most crimes are reported in the Central District, followed by Northern and Southern. Please

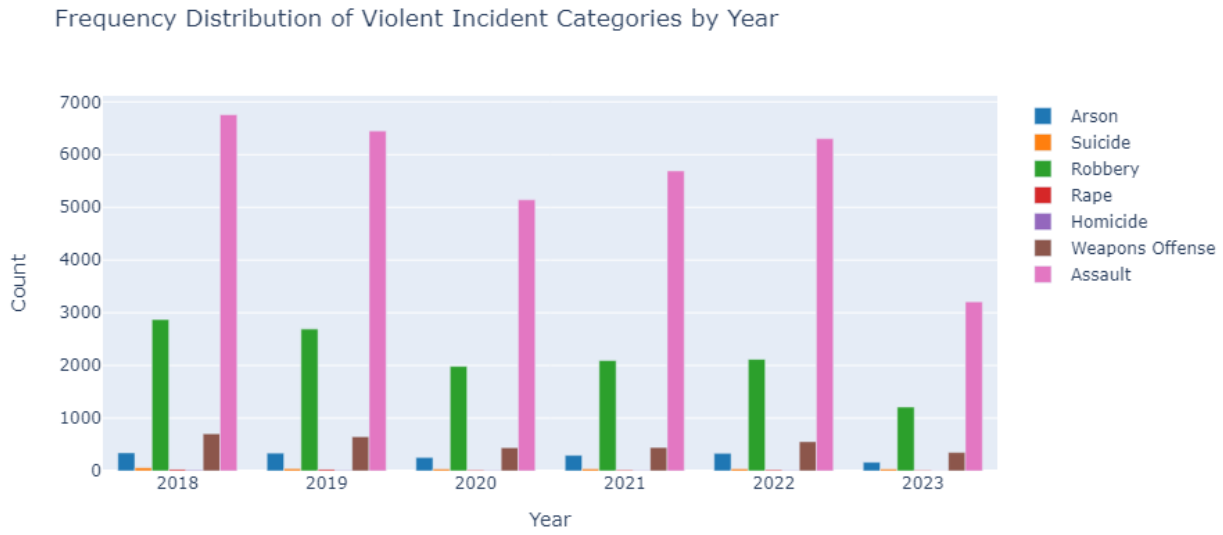
note that the number of incidents in 2023 appears to be half of the usual number because this analysis is done in July 2023 and we have only 6 months data from 2023.

Insight 6: Top 10 Crime Categories



From the above Pie Chart, we can see that Larceny Theft seems to be the most committed crime, covering over 41% of all crimes recorded in the last 4 years. It is followed by Malicious Mischief and other incidents, which cumulatively account to only 16%.

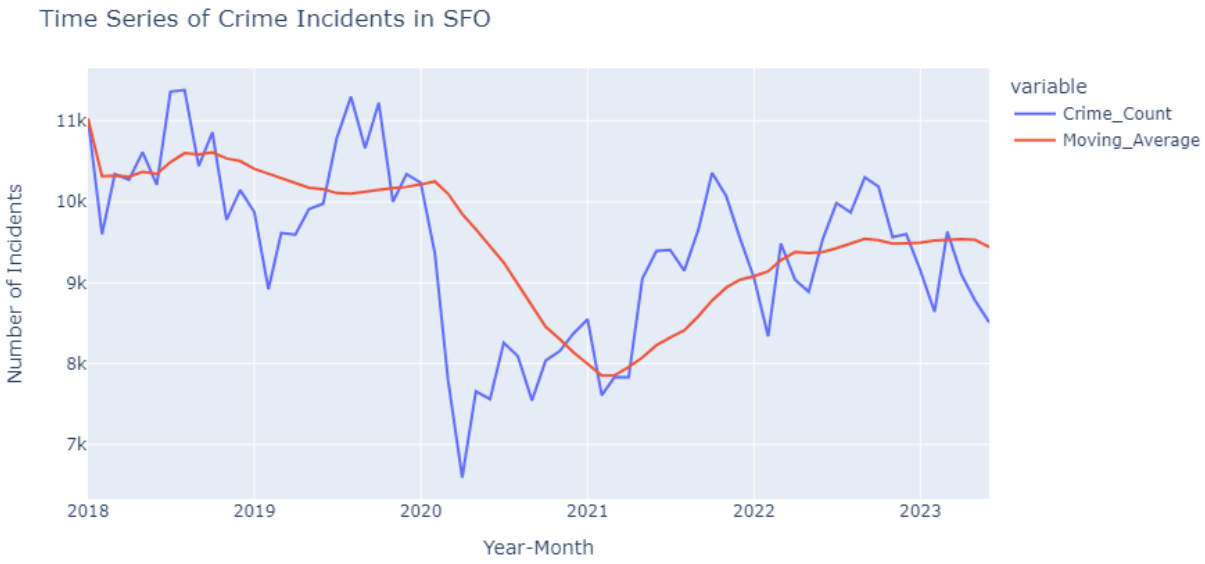
Insight 7: Year-wise Crime Counts for Violent Crimes



From the above grouped bar chart, we can observe that over the years, Assault has been the most prevalent violent crime in San Francisco, followed by Robbery. We can notice that highly violent crimes like Rape, Murder, etc. are not very high in number.

Time Series Analysis

We perform a time series analysis of all the crimes reported in San Francisco from 2018-2022 to identify if there is any pattern or trend. Below is the graph that visualizes the number of incidents reported per year and the moving average. We can observe a significant dip in the number of crimes reported in 2020, which might be attributed to COVID-19 pandemic. Post the pandemic, when everything slowly resumed to normal, we can also see that crime rate also started to pickup.



Recommendations

Based on the above key insights from exploratory data analysis of crimes in San Francisco, here are some recommendations that can be considered:

1. **Targeted Policing Efforts**: Considering the higher number of crime incidents in January, June, and May, it would be beneficial to allocate additional resources and increase patrols during these months to proactively address potential crime spikes.
2. **Enhanced Friday Patrols**: Given that Fridays have higher crime incidents, focusing law enforcement efforts on this day of the week can help deter criminal activities and ensure a safer environment for residents and visitors.
3. **Daytime Security Measures**: Since crime incidents are significantly higher during the daytime, particularly from mid-day to 6 PM, it would be valuable to enhance security measures, such as increased police presence, in areas with higher foot traffic during these hours.

4. **Encourage Online Crime Reporting**: While most crimes are reported offline, there may be an opportunity to promote and facilitate online crime reporting mechanisms. This can help streamline the reporting process, improve data accuracy, and ensure timely responses to incidents.
5. **Concentrated Efforts in High-Crime Districts**: Given that the Central District, followed by the Northern and Southern districts, have the highest number of reported crimes, allocating additional resources and implementing targeted crime prevention strategies in these areas can help address the concentration of criminal incidents.
6. **Addressing Larceny Theft**: With Larceny Theft being the most committed crime, accounting for over 41% of all recorded crimes, it is crucial to implement preventive measures such as increasing surveillance, educating the public about safeguarding their belongings, and enforcing stricter penalties for theft-related offenses.
7. **Focus on Reducing Assault and Robbery**: Since Assault and Robbery are the most prevalent violent crimes in San Francisco, targeted initiatives aimed at prevention, community engagement, and early intervention can contribute to reducing these types of incidents. Collaborating with local organizations and implementing community-oriented policing programs can help foster trust and proactive crime prevention.

Conclusion

In conclusion, the data analysis of crimes in San Francisco has provided valuable insights into the patterns and trends within the city's criminal landscape. By examining various aspects such as the temporal distribution of crime incidents, the day of the week with higher occurrences, crime reporting methods, district-wise crime concentrations, and the prevalence of specific crime types, we have gained a deeper understanding of the dynamics of crime in the city.

Furthermore, the findings highlight the importance of promoting online crime reporting mechanisms to improve data accuracy and expedite responses. Concentrated efforts in districts with higher crime rates, particularly in addressing the prevalence of Larceny Theft, can help reduce overall crime incidents and enhance community security.

It is crucial to acknowledge that the fight against crime is an ongoing process that requires continuous evaluation, collaboration between law enforcement agencies, community involvement, and adaptation to evolving crime patterns. By leveraging data-driven insights and implementing targeted interventions, we can collectively work towards a more secure and thriving San Francisco.

This project underscores the importance of utilizing data analysis as a powerful tool in understanding crime patterns, informing decision-making, and ultimately fostering a safer city for all. It is my hope that the findings presented here contribute to the ongoing efforts of the police department and stakeholders in creating a secure and inclusive community in San Francisco.