# Product Intern Assignment

## **Assignment:** Data Analytics of Time Series Data

**Objective:**
You have been tasked with a project to uncover insights from a time series dataset. This project will challenge your research skills to find a suitable dataset, your analytical thinking to ask pertinent questions, your data analytics skills to answer these questions, and finally, your documentation and presentation skills to convey your findings effectively.

Dataset-https://www.kaggle.com/datasets/vaghefi/santa-clara-country-crime

Dataset column count-20

Dataset row count-194865

Ipynb link- click here

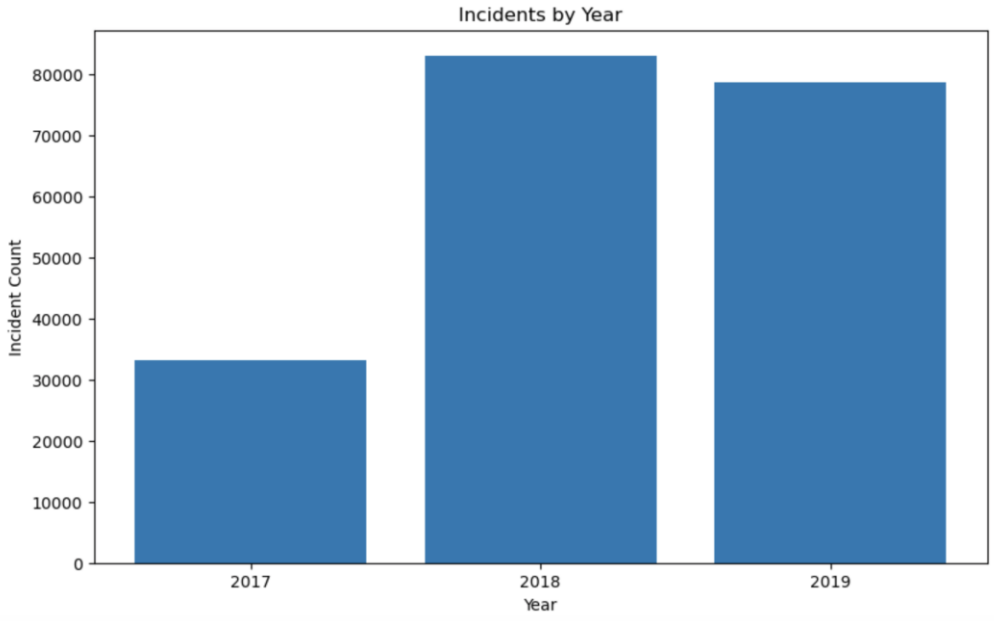
Content-The dataset contains all the crime reported in Santa Clara County, CA from Aug 2017 to Dec 2019. Crime data reported from 911 calls in Santa Clara County, California, USA.

Github link-

# Report

Q1) **What are the overall crime trends in the city over the past few years?**

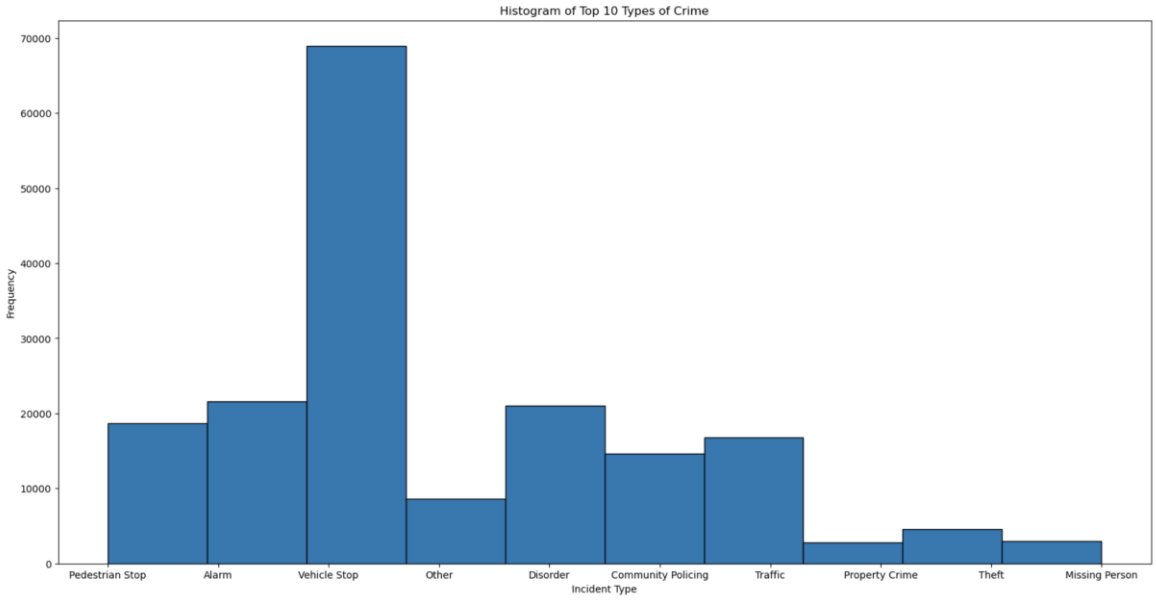
The trend in crime rates for the years 2017, 2018, and 2019 appears to show some fluctuations, albeit within a relatively narrow range. In 2017, the mean crime rate was recorded at 6,628 incidents, which then increased slightly to 6,925 incidents in 2018, before decreasing to 6,551 incidents in 2019. This pattern suggests that there was an initial uptick in crime from 2017 to 2018, followed by a subsequent decline in 2019. While the data indicates a modest rise and fall in crime rates during this period, it's important to note that these numbers do not provide the full context. Factors such as changes in law enforcement strategies, socioeconomic conditions, and shifts in population demographics can all influence crime rates. A more comprehensive analysis would be needed to determine the underlying causes of these fluctuations and whether they represent a meaningful long-term trend.




The trend in crime rates for the years 2017, 2018, and 2019 reveals interesting insights when considering the context that the data for 2017 covers only a 5-month period.

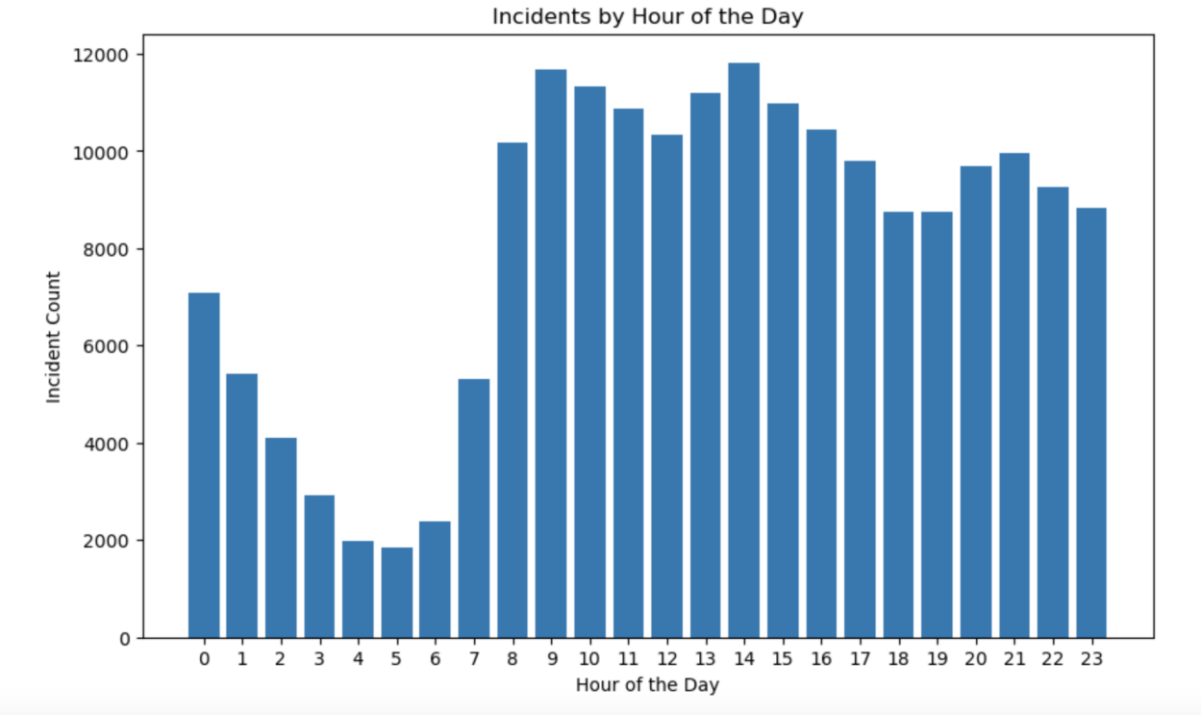
Q2) **What are the most common types of crimes in the city?**

The ten most reported crime categories, including pedestrian stops, alarms, vehicle-related incidents, disorders, community policing matters, traffic-related property crimes, thefts, and missing person cases, offer a comprehensive snapshot of law enforcement concerns in a given area. Each category represents a unique facet of public safety and community well-being. Pedestrian stops indicate proactive policing efforts, while alarms address security concerns. Vehicle-related incidents encompass various scenarios, from accidents to thefts, emphasizing the significance of road safety. Disorders and community policing underscore the importance of maintaining social order and community engagement. Traffic-related property crimes, thefts, and missing person cases shed light on property-related offenses and safety concerns. Together, these crime categories reflect the diverse challenges that law enforcement agencies address to ensure the security and welfare of their communities. Analysing and responding to these reports effectively is crucial for maintaining public safety and fostering community trust in law enforcement agencies.




Q3) What is the time of day when crimes are most frequently reported?

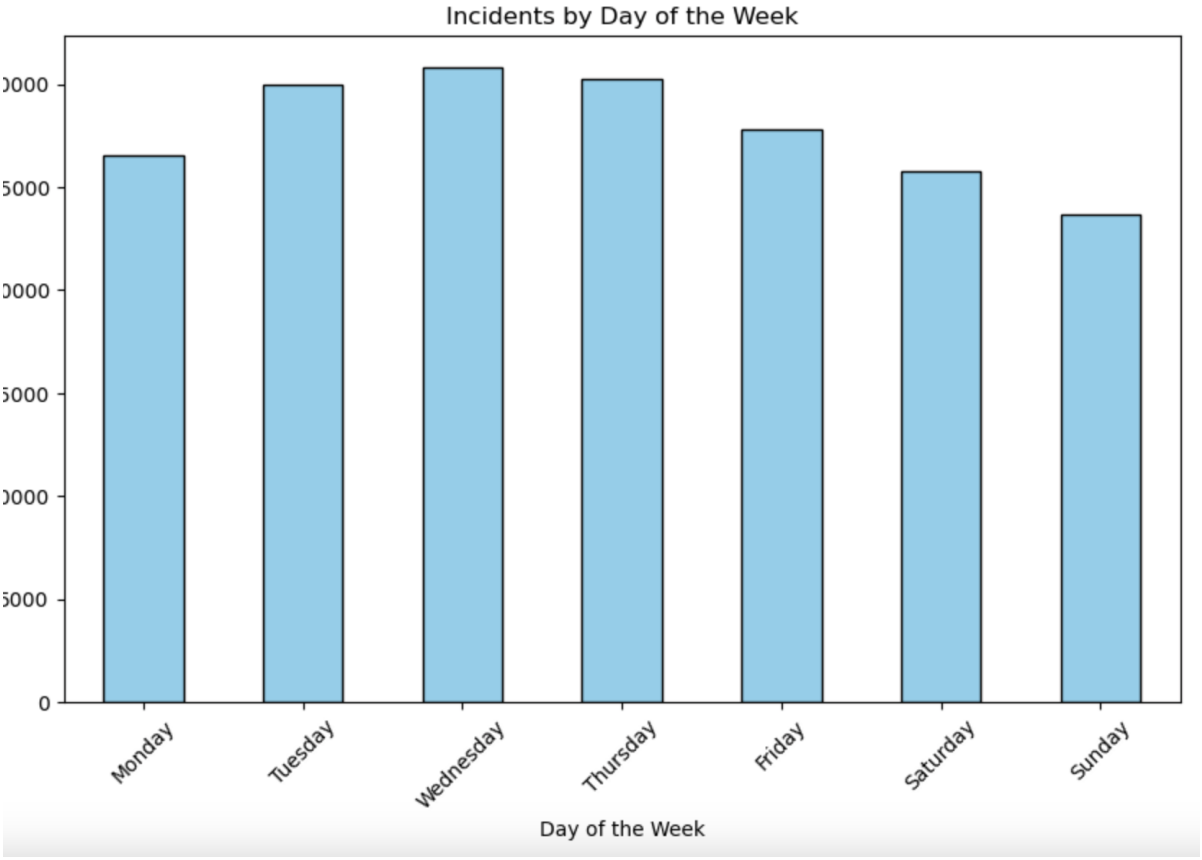
The analysis of crime patterns throughout a 24-hour period reveals intriguing insights. Crime rates exhibit a notable decrease during the early morning hours, particularly between 1 to 7 AM. However, as the day progresses, from 8 AM to 4 PM (16:00), there is a noticeable uptick in reported incidents. This period corresponds with the typical daytime activities of businesses, schools, and workplaces, potentially contributing to a higher visibility of criminal behaviour. Interestingly, after this peak, there's a gradual decline in crime rates as the evening progresses, albeit slightly, leading up to midnight. This fluctuation in crime rates throughout the day underscores the importance of considering the temporal aspect in law enforcement strategies and community safety efforts, with particular attention to peak hours for effective crime prevention and management.




Q4) Is there any trend in the occurrence of crimes during the week?

The analysis of crime data highlights a distinctive pattern: a higher frequency of incidents on weekdays compared to weekends, particularly during non-holiday periods. This phenomenon can be attributed to the routine activities that typically

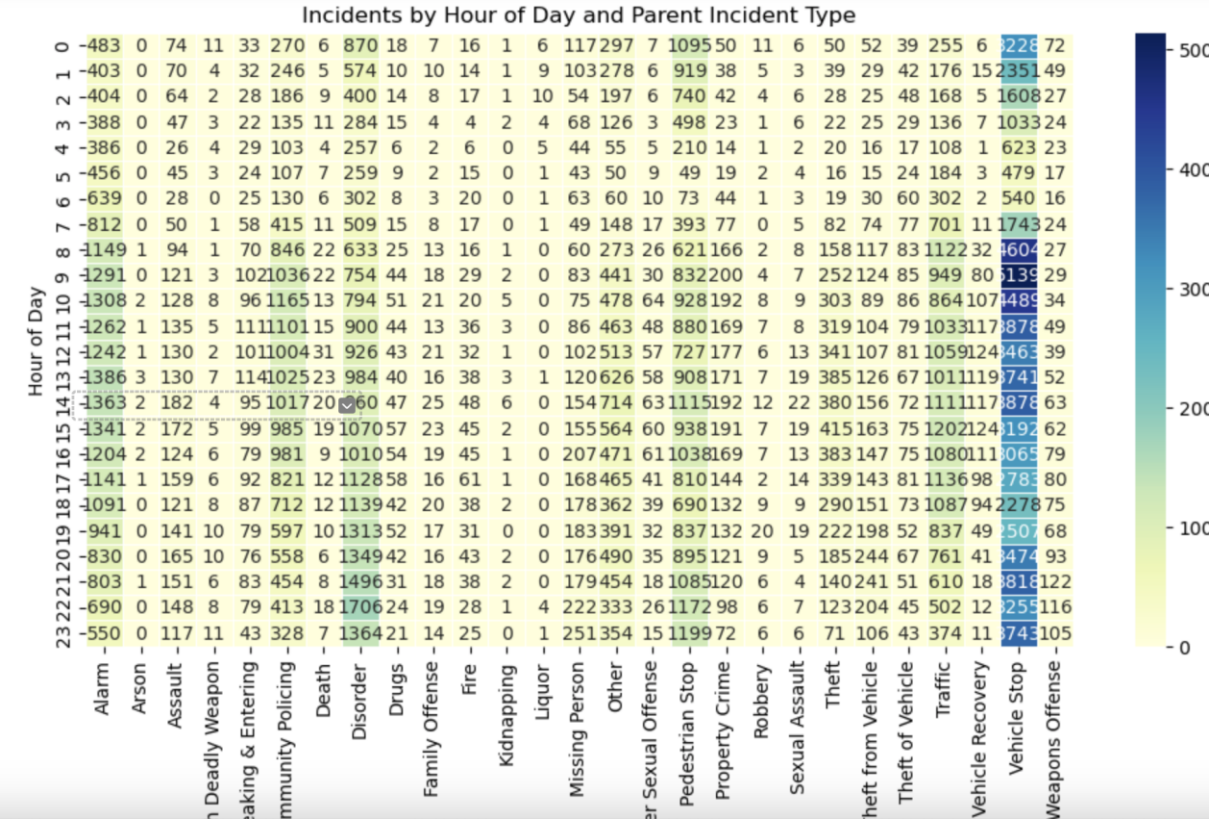
fill weekdays, such as work, school, and business operations, creating a more conducive environment for certain types of crimes. In contrast, weekends, especially during holidays when many people have time off, tend to see a decrease in criminal incidents. This shift in crime rates underscores the complex relationship between daily routines, opportunities for criminal activity, and the impact of holidays on community safety dynamics. It also emphasizes the need for law enforcement agencies to adapt their strategies to reflect these fluctuations and prioritize resources accordingly to maintain public safety.




Q5) Are there specific time-of-day patterns in crime occurrence?

The heatmap illustrating incidents by the hour of the day and parent incident type provides valuable insights into the temporal distribution of crime. One prominent observation is the concentration of criminal incidents between the hours of 08:00 AM to 04:00 PM (16:00). During these daytime hours, various types of crimes tend to be more prevalent. This phenomenon aligns with the hustle and bustle of daily life, as well as increased opportunities for criminal activities in commercial

and residential areas. Understanding this peak in crime rates during the daytime allows law enforcement agencies to strategically allocate resources and implement preventive measures to address and mitigate criminal activities during these hours. It also highlights the significance of considering the temporal dimension in crime analysis and public safety efforts.




Q6) What does the analysis of incident date-time and updated_at time data suggest about police office?

From the available incident date-time and updated_at time data, it appears that the Santa Clara County police have demonstrated an efficient response and resolution process. It can be observed that each incident, based on the timestamps of incident date-time and updated_at, is typically resolved within a relatively short timeframe of 7 days or less. This prompt resolution of incidents reflects the dedication and effectiveness of law enforcement in the region.

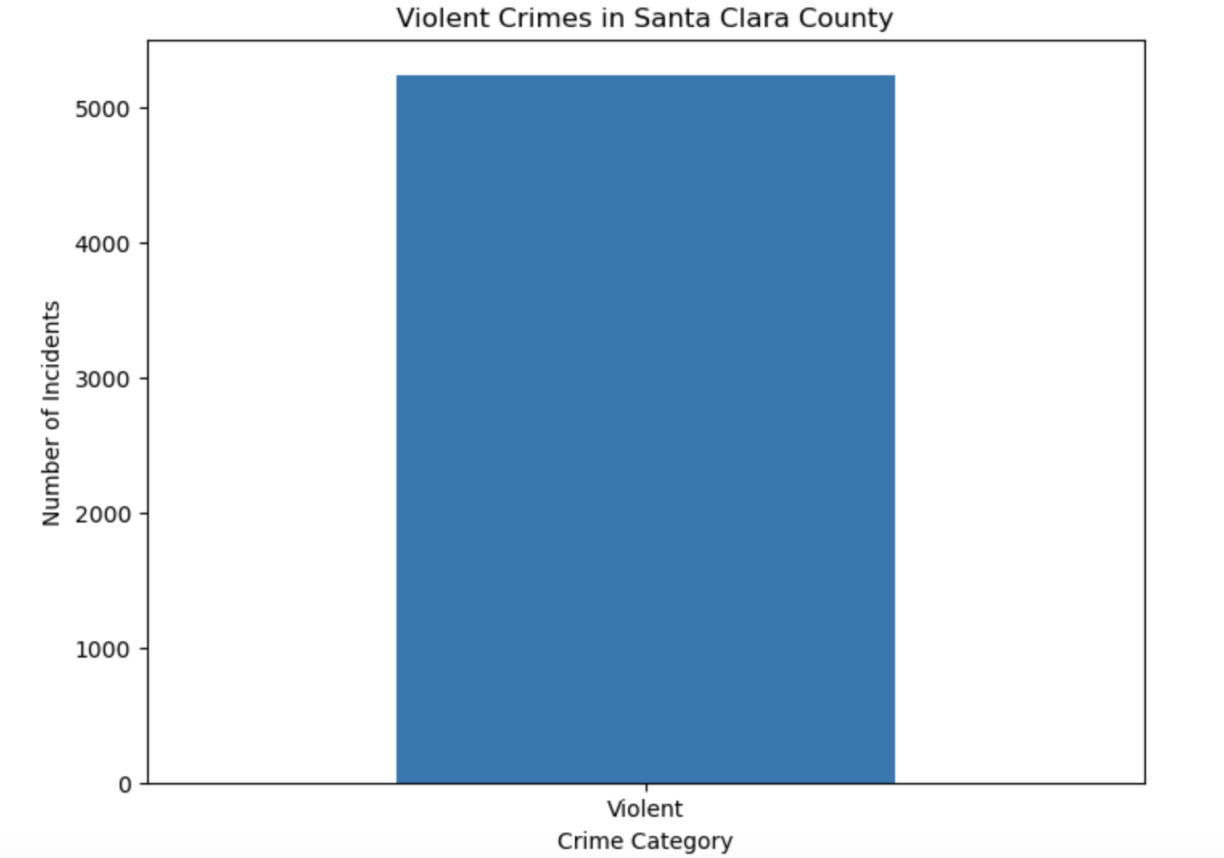
In [51]:
```
# Convert 'incident_datetime' and 'updated_at' columns to datetime objects
df['incident_datetime'] = pd.to_datetime(df['incident_datetime'])
df['updated_at'] = pd.to_datetime(df['updated_at'])

# Calculate the day gap between 'updated_at' and 'incident_datetime'
df['day_gap'] = (df['updated_at'] - df['incident_datetime']).dt.days

# Display the DataFrame with the day gap
df['day_gap']
```

Out[51]:
```
0         7
1         7
2         7
3         7
4         7
         ..
194860    7
194861    7
194862    7
194863    7
194864    7
Name: day_gap, Length: 194865, dtype: int64
```

Q7) Is there a significant prevalence of violent crime in Santa Clara County?

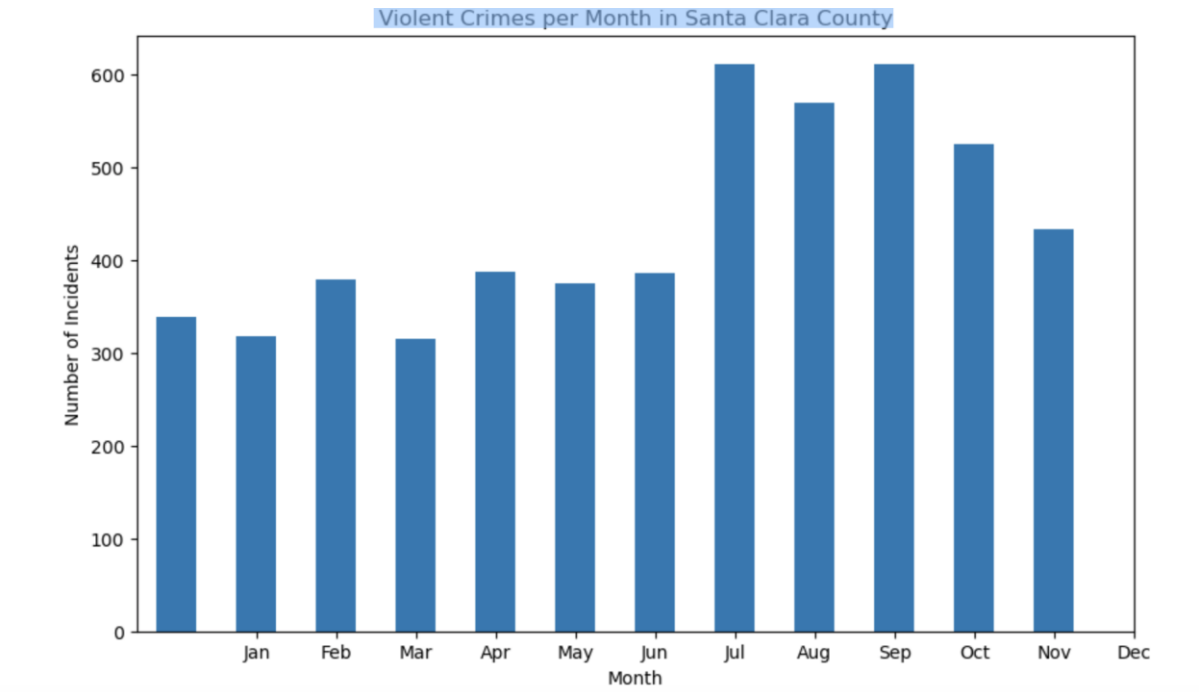
Violent crimes, encompassing categories such as Assault, Weapons Offenses, Other Sexual Offenses, Sexual Assault, Assault with a Deadly Weapon, Kidnapping, Arson, and Robbery, often garner significant attention due to their immediate impact on victims and communities. However, in the context of the broader crime landscape, it's noteworthy that non-violent crimes significantly outnumber their violent counterparts, with a staggering count of approximately 189,000 incidents. Non-violent crimes, which include various property crimes,

thefts, and related offenses, while less physically harmful, can still have substantial financial and emotional repercussions for victims and society. The coexistence of both violent and non-violent crimes underscores the multifaceted nature of law enforcement's responsibilities, necessitating a balanced approach to address the diverse challenges posed by different types of criminal activities.




Q8) What is the month wise distribution of violent crime in Santa Clara County?

The incidence of violent crime in Santa Clara County typically hovers around 350 to 400 incidents per month, reflecting a relatively stable baseline throughout the year. However, a noticeable shift occurs during the summer season. During these warmer months, there is a discernible increase in violent crime rates. This seasonal variation suggests that factors such as increased outdoor activities, higher temperatures, and potential changes in social dynamics may contribute to the uptick in violent incidents. Law enforcement agencies and community stakeholders often pay close attention to these patterns to ensure adequate resources and strategies are in place to address and mitigate the challenges associated with the seasonal rise in violent crime.




Q9) Is the primary incident reported and the actual incident reported same?

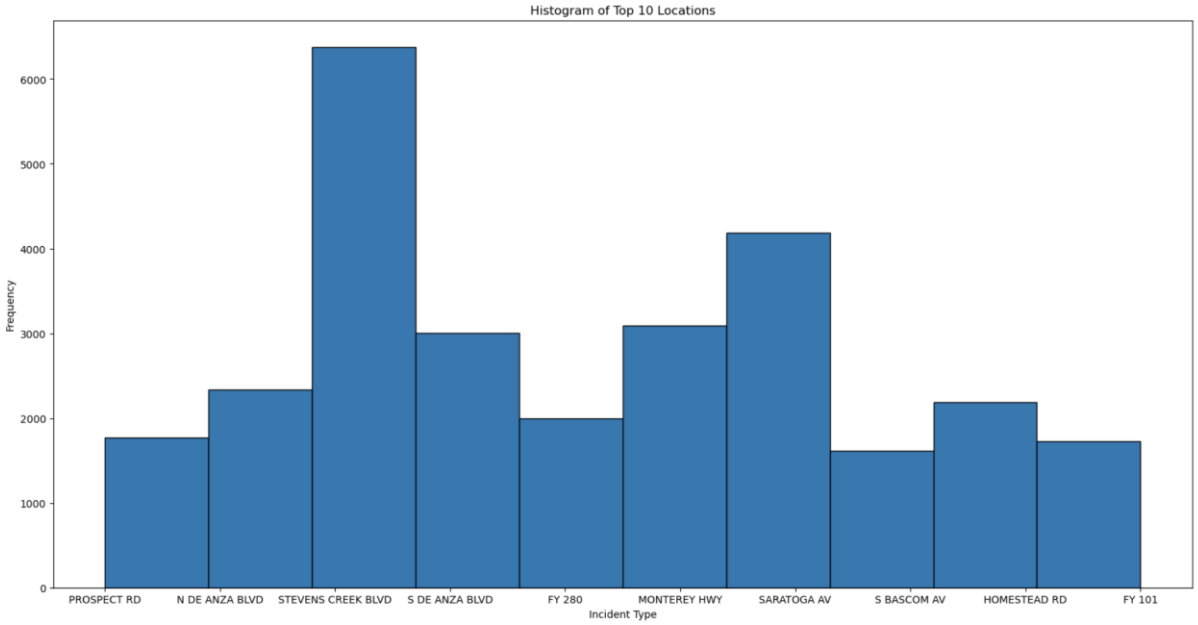
The data indicates a significant disparity between the number of primary incidents reported and the actual incidents reported. Specifically, there are 8,603 cases where the primary incident type matches the actual incident type, while a substantially larger number, 186,262 cases, show a discrepancy between the two. This difference underscores the complexity of incident reporting and classification, revealing that in a considerable majority of cases, the primary incident type does not align with the detailed nature of the events reported. Such disparities may result from varying reporting standards or the multifaceted nature of certain incidents that may encompass multiple elements. Understanding and addressing these differences are essential for accurate crime analysis and effective law enforcement strategies.

```
df['Same_Type'] = (df['incident_type_primary'] == df['parent_incident_type'])

same_type_counts = df['Same_Type'].value_counts()

# Print the counts
print(same_type_counts)
```

```
False    186262
True       8603
Name: Same_Type, dtype: int64
```

Q10) What are the common locations of crimes in the city?

The diagram provides valuable insights into the top 10 locations where the highest number of incidents have occurred. At the forefront, Stevens Creek claims the number one position, followed by Saratoga Avenue at the second spot, and Monterey Highway securing the third rank. These rankings shed light on areas within Santa Clara County that warrant particular attention from law enforcement agencies and community stakeholders. Understanding which locations are more susceptible to incidents can help in allocating resources strategically, implementing targeted safety measures, and fostering proactive engagement with local communities to enhance security and well-being in these areas. By addressing the unique challenges posed by these top-ranking locations, law enforcement can work towards creating safer environments for residents and visitors alike.




Name-Darshan Dipak Patel

Email-bedarshanpatel@gmail.com

Phone No.-9993744626