
**Final Project – Police Incidents Reported in San Francisco 2018 - 2022**

ALY 6070: Communication and Visualization for Data Analytics

Assignment Completion Date: 05/18/2022

Instructor: Professor Jack Bergersen

Term: Spring 2022

Phimonkae Wilairat

**Introduction:**

According to the think tank Hoover Institution, San Francisco is a dangerous city to live in based upon crime rate compared to other cities in the United States. Statistics provided by the Hoover Institute show that residents who live in San Francisco face a 6.25% chance of being a victim of a property or violent crime (Ohanian, L., 2021). There are many factors that lead to the high rate of crime in the city, and one way to begin analyzing these factors is reviewing the Police Department Incident Reports dataset from 2018 to 2022 provided by the San Francisco Police Department (SFPD) compiled within the department's Crime Data Warehouse (CDW) (SFPD, 2021). Some of the factors that we will analyze from this dataset include the count, location, and type of incidents reported by police district and neighborhood within the city. We will analyze these factors by providing effective visualizations created using Tableau. Additionally, we will provide insights into our design decisions for each visualization.

The type of incident reports to be analyzed:

1. Initial Reports: the initial or first report filed for a crime incident
2. Supplemental Reports: Follow ups to the first report filed
3. Coplogic Reports: Reports created by the public using the San Francisco Police Department's online reporting system
4. Vehicle Reports: crime incident reports regarding stolen vehicles (SFPD, 2021)

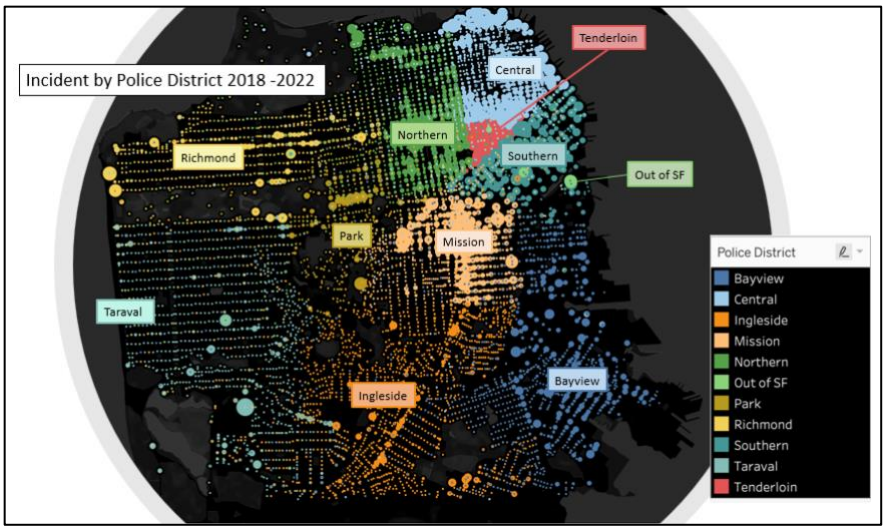
We will analyze the crime incident reports by using the "Where, When, What, and How" methodology. Additionally, we will split the analysis into three parts:

1. How are incidents reported by location (police district) and incident category?
2. What times do crimes occur?
3. How are all incidents recorded?, How do variables (such as neighborhood) affect the number of reports generated?

**Part I: How are incidents reported by location (police district) and incident category?**

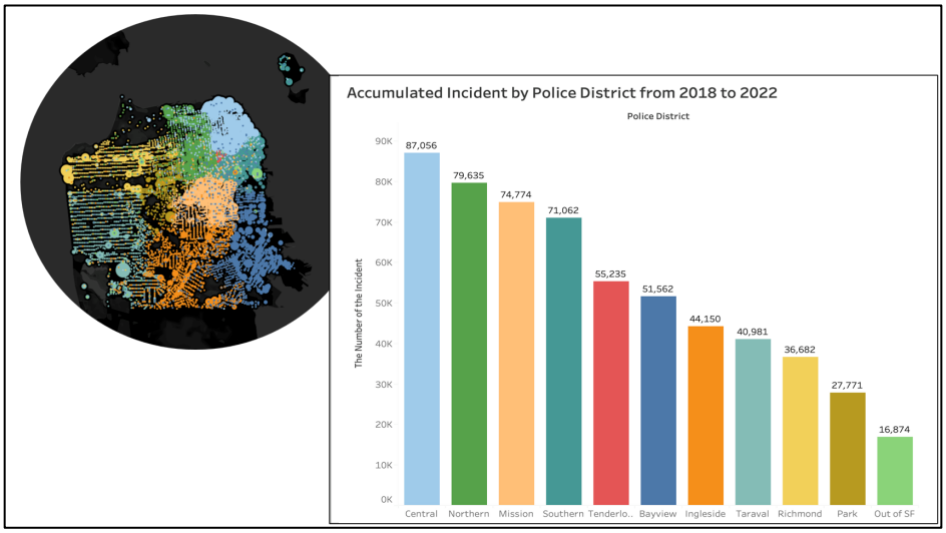
To begin understanding the crime that occurs in San Francisco, we will begin by analyzing the location in which crime occurs, and the types of crimes that occur. We will answer the following research question by focusing on the following points:

1. Understand which area has the most incidents reported.
2. Provide details of what kind of incident occurred in each area the most.
3. Understand the change in the number of reports made by year in each area. We plan to extract data in the criminal report to see the trend of criminal incidents and whether they increase or decrease over time.



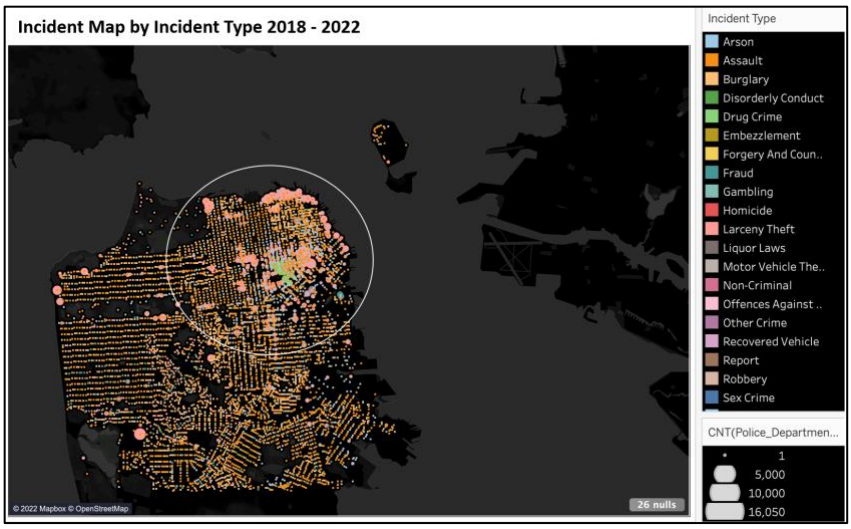

*Plot 1: Incident by Police District 2018 to 2022*

The visualization in plot 1 shows a map that depicts the count of police incidents by district from 2018 – 2022. The color of each bubble represents which district the incident occurred in. The size of each bubble represents the count of incidents over the given time-period. By analyzing this map, the audience can quickly identify that the most concentrated region of incidents is towards the upper-right of the map, encompassing the Tenderloin, Central, Northern, Southern police districts. Additionally, the Bayview district appears to have larger circles compared to other areas in the map, signifying higher counts of incidents.



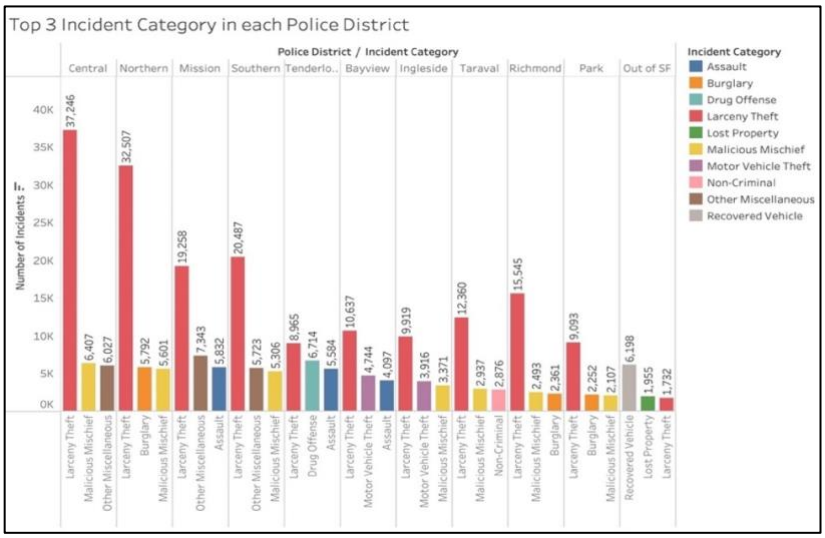

*Plot 2: Accumulated Incident by Police District from 2018 to 2022*

The visualization in plot 2 shows a bar chart with the total number of incidents by police district from 2018 to 2022. The color coordination of the bars corresponds to the map in the upper left-hand corner: The color represents the police district in which the incident occurred. As described in the explanation for plot 1, the police districts with the highest amount of crime incidents recorded is confirmed with the bar chart visualization.



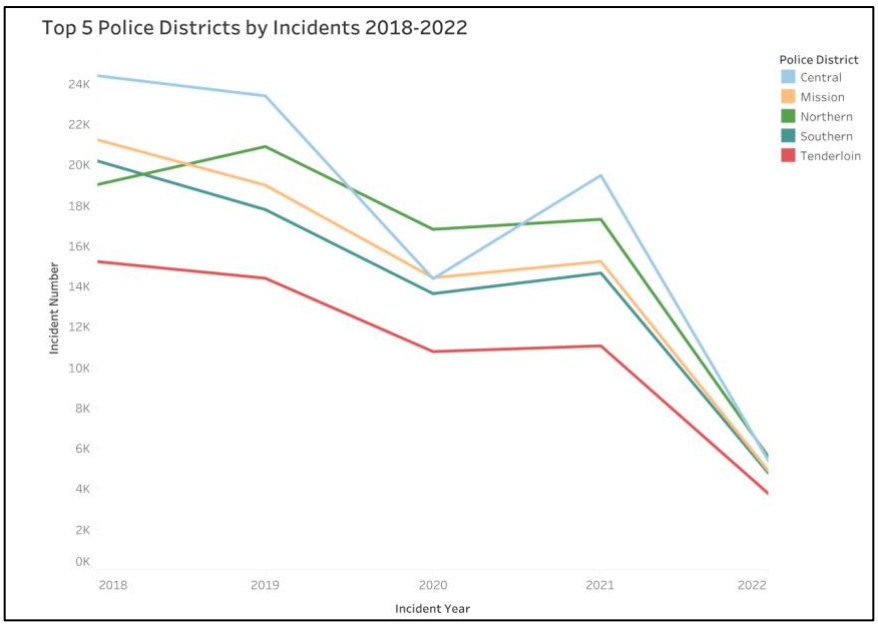

*Plot 3: Incident Map by Incident Type 2018 – 2022*

The visualization in plot 3 maps the incident location by incident type. This allows the audience to review what type of incidents occur in what part of the city. The color of the bubbles represents the type of crime. The size of the bubbles represents how large the count of incidents is for each type of crime. By analyzing this visualization, the audience can see that the larceny theft represents the most frequent incident that occurs, especially in the top-right region of the map. Additionally, the multiple bar chart in plot 4 below reinforces that larceny theft has the highest count of incidents, in every police district. We chose the top 3 most prevalent incident types to show the relationship between each police district. Each color bar represents a different type of incident, while each 'block' represents a different police district in the city. When consider this Plot 4 with Plot 3, Central, Northern, Mission, Southern, and Tenderloin, all of them located in the top-right area. We can see that there is the concentration of a pink bubbles that represent the larceny in these areas. Interestingly, drug crime occurs in Tenderloin have a significant number, second only to larceny theft. We can notice the mint color obviously in the map, Plot 3.




*Plot 4: Top 3 Incident Category in Each Police District*

Finally, we conclude Part 1 by analyzing the top 5 police districts ranked by number of crime incidents from 2018 to 2022.




*Plot 5: Top 3 Incident Category in each Police District*

The time-series line chart above shows the trend of overall crime incidents throughout the last four years. We chose only the top five police districts, as including more districts would make the chart too busy and confuse the audience, preventing them from making any conclusions from the visualization. By looking at this visualization, it is clear that the amount of overall incidents for each district is trending downwards. It is likely that the overall amount of incidents has reduced due to the COVID-19 pandemic. According to Tom Jackman of the Washington Post, may large cities in the United States experienced significant drops in crime during the pandemic. This may be attributed to fewer people out in the streets of the city, and primarily staying in their homes during this time-period (Jackman, T., 2020).

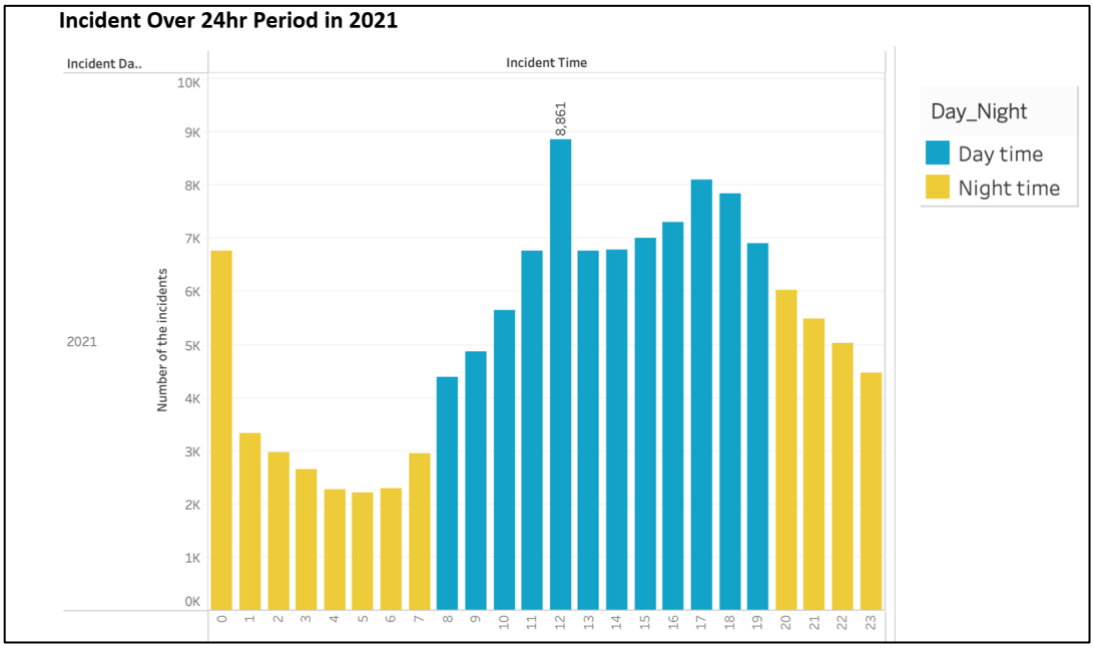
**Part 2 : What times do crimes occur?**

In addition to the location and type of incidents that occur, it is important to understand the time that incidents occur so that more police force can be staffed at these times and the public of San Francisco can be more alert. We will give the audience a better understanding of the following points through data visualizations:

1. When the most reported crimes occurred within a 24hr. period.
2. When does each specific type of crime usually occur (day-time vs night-time).

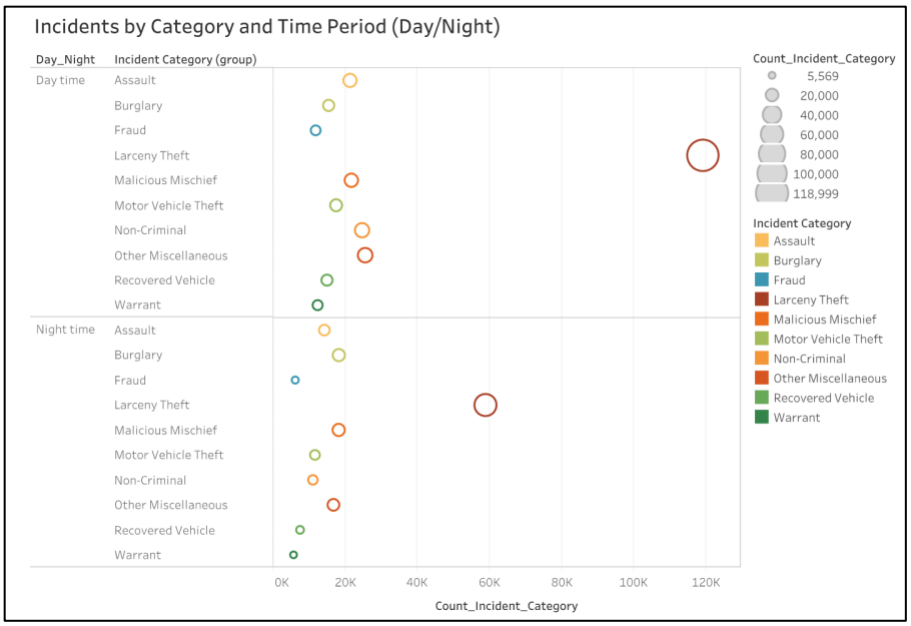
As visualized on the next page, plot 6 shows the incidents that occurred during each hour of the day during 2021. Rather than show a visualization for each year covered in this analysis, we presented the most recent full-year, 2021, as the trend was consistent throughout the four-year

period from 2018 to 2022. The chart below shows the accumulated amount of incidents in each hour. Blue represents day-time hours, yellow represents night-time hours. The distinction of colors clearly shows that the majority of incidents occur in the day-time compared to the night-time hours. Furthermore, the incidents are at their lowest point between the hours of 1AM and (hour 1) and 7AM (hour 7). We chose to include gridlines to provide the audience with a clear reference point to the amount of incidents occurred, as we displayed only one data point, the hour with the most accumulated incidents occurred (The 12PM hour).



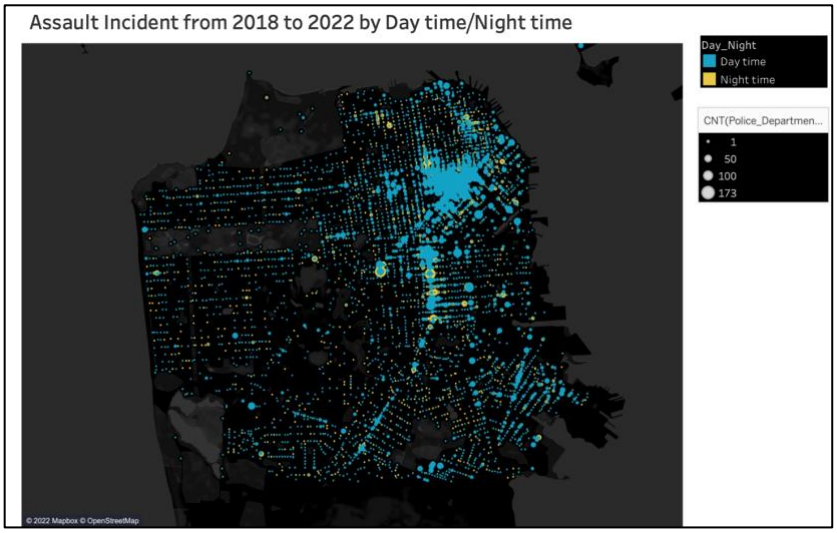

*Plot 6: Incidents that Occurred During Each Hour of the Day During 2021*

In plot 7 below, we displayed the type of incidents occurred during the day-time compared to the night-time. Each circle is bordered with a different color to signify incident category. The size of the circle represents the amount of incidents that have occurred for each incident. The two sections, top and bottom, represent day-time (hours 0-12) and night-time (hours 13 – 24), respectively. This visualization shows two large outlying circles to the rights of the chart which represent larceny theft. This shows that compared to other crimes, larceny occurs much more frequently in the day-time compared to the night-time. It seems counter-intuitive, but more thief do their misdeeds in the middle of the day instead of under cover of darkness when we consider plot 7 together with plot 6. From both graph, the larceny victimizations are revealed that the most common time was between 12 PM to 18 PM. However, It appears that the other crimes have similar incident counts in both the day-time and night-time.




*Plot 7: Incidents by Category and Time Period*

The final visualization of Part II presents a map (plot 8) of assault incidents that represents those that occur in the day-time versus those that occur in the night-time from 2018 through 2022. The crime incidents that occur in the day-time are represented in blue bubbles, while the incidents that occur in the night-time are represented by yellow bubbles. The size of the bubbles represents the count of assault incident by location. We chose to display assault incidents because they are a top 5 occurring incident by incident category, and the map shows clearly areas that have high incidents that occur in the day-time versus those in the night-time. Our analysis shows that day-time assaults are concentrated towards the upper right area of the map, although there are specific areas that have been susceptible to night-time assaults, depicted by larger yellow bubbles.

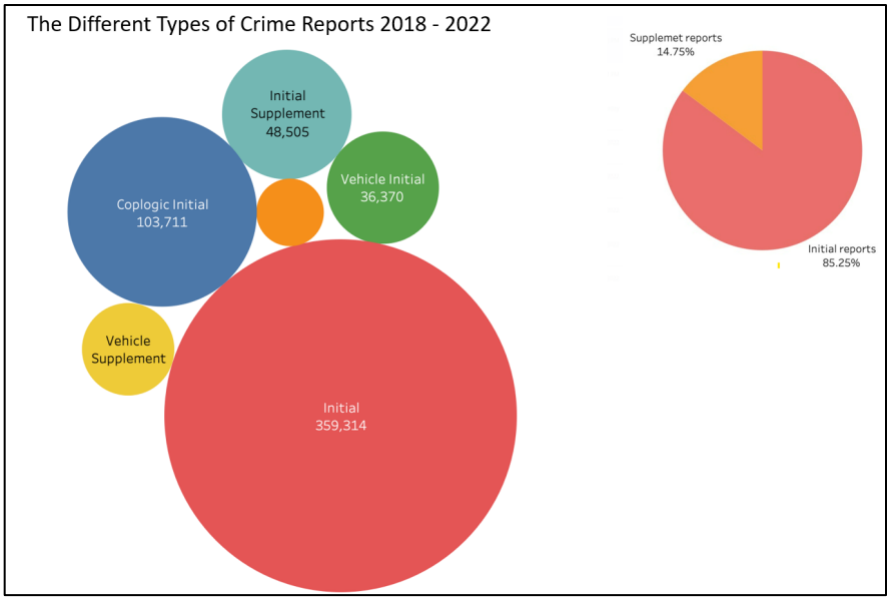



*Plot 8: Assault Incidents from 2018 to 2022 by Day-Time / Night-Time*

**Part 3: How are all incidents recorded? How do variables (such as neighborhood) affect the number of reports generated?**

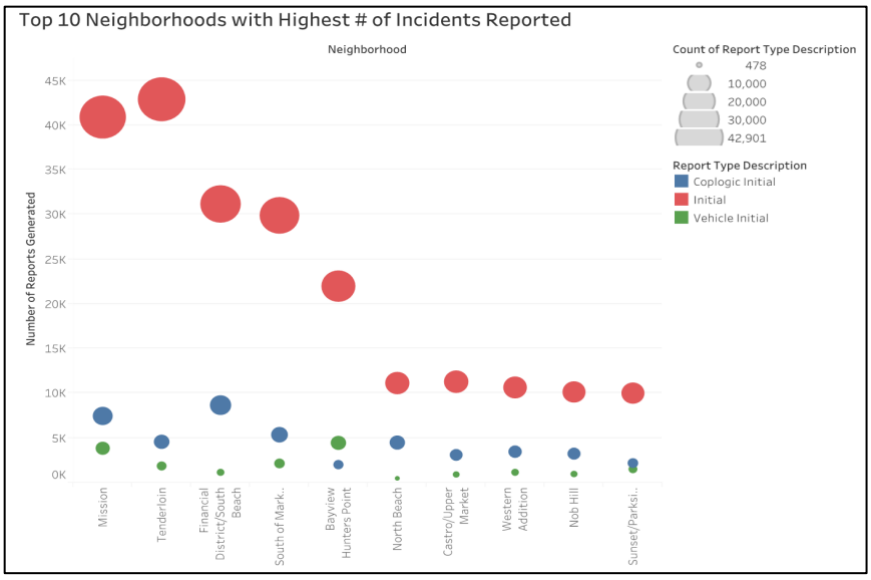
In the final part of our analysis, we have uncovered insights based upon how incidents are recorded, the type of reports that are reported, and the incidents ultimate outcome using various types of visualizations.

1. Utilize a bubble chart and pie chart to show the amount of crime incidents reported from 2018 to 2022.
2. Use bubble combination chart to show the highest crime incidents reported by neighborhood.
3. Use a line graph to show the trend in crime incident reported over time by report type.



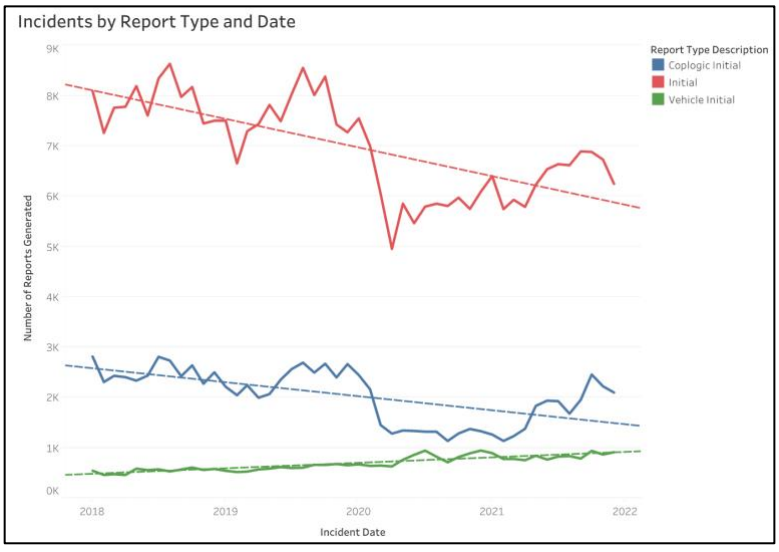

*Plot 9: Different Types of Crime Reports 2018 to 2022*

In plot 9 above, we display a bubble chart that visualizes the count of each report type from 2018 to 2022 by the size of the bubble. The larger the bubble, the higher the count of the respective report type. We are displaying both the initial and supplement reports. From plot 9, we can observe that there are six categories of reports under which the incidents are recorded. Highest number of incidents are recorded under Initial reports. Three categories of reports are initial type of reports and the other three are supplement reports. Hence, we wanted to analyze how many are initial reports and how much percent of reports are supplement which are just additional reports of further investigations. Hence, we have included a pie chart in the upper right-hand corner that displays the percentage of overall initial reports versus supplemental reports. From the pie chart we observed that almost 85% of the reports are initial reports and only 15% belongs to supplement reports. Hence, we have considered only initial reports in further analysis to study the impact of other attributes on the number of reports generated.




*Plot 10: Top 10 Neighborhoods with the Highest # of Incidents Reported 2018 to 2022*

Plot 10 presents a bubble combination chart the shows the top 10 neighborhoods based upon the highest number of crime incidents reported from 2018 to 2022. We describe this visualization as a "bubble combination" chart because it is combines components of a bar chart and a bubble chart. The chart is in descending order from neighborhoods with the highest number incident to neighborhoods with the least number of incidents. The size of each bubble represents the count of each report type. The color scheme corresponds to the last visualization (initial reports are red, coplogic reports are blue, vehicle initial reports are green). Visually, the Tenderloin neighborhood has a higher Initial report bubble than the Mission neighborhood, however the overall total count is less than Mission, which is why Tenderloin is ordered as the second neighborhood in the visualization.

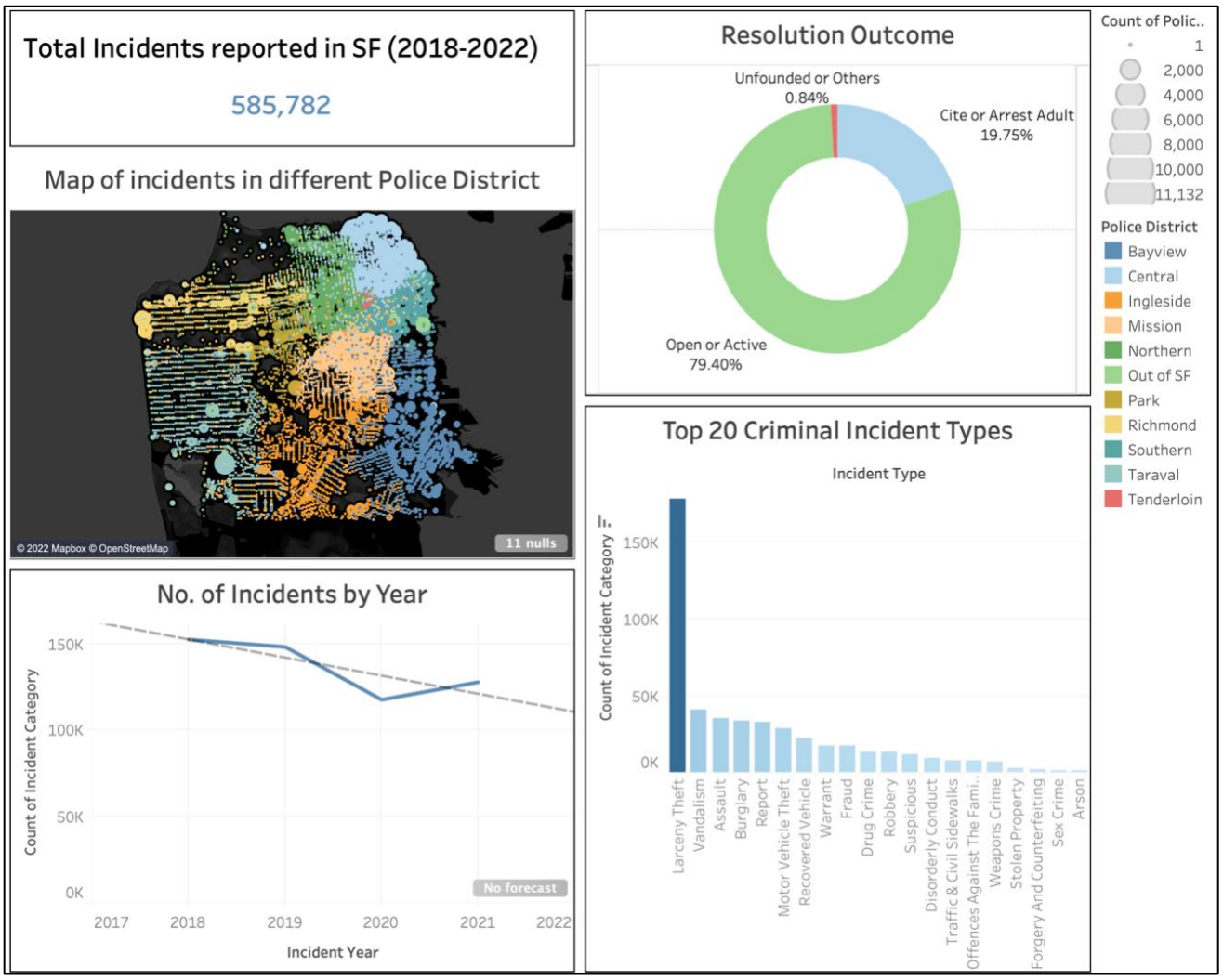



*Plot 11: Top 10 Neighborhoods with the Highest # of Incidents Reported 2018 to 2022*

Plot 11 above shows a time-series line graph of the initial criminal incident reports from 2018 to 2022, excluding supplemental reports. We have included a regression trend line to clearly show the audience the trend of incidents over the period reported. The colors are consistent with the previous two visualizations. From an analysis perspective, it is interesting to note that Initial reports had the biggest drop in reports during 2020, likely driven by the pandemic, while initial vehicle reports had a slightly positive trend, with very little drop in 2020.

## Dashboard

In addition to our individual visualizations, we created a Tableau dashboard to highlight our overall analysis and provide high-level guidance to our audience in one package.




*Plot12:Dashboard*

In our final visualization, the dashboard above summarizes total incidents reported in San Francisco form 2018 to 2022, a map of incidents by police district with corresponding legend to

the right of the dashboard, a time-series displaying the total number of incidents per year with regression trend line, a resolution outcome donut chart, and a bar chart the displays the top incident types throughout 2018 to 2022. We decided that using a variety of visualization types would keep the audience interested in reviewing this visualization, while eliminating any extraneous clutter. From a visualization creation perspective, we spent the most time developing the Resolution Outcome donut. Although simple in appearance, the creation of the visualization was complicated. First, we grouped two categories, 'Exceptional Adult' and 'Unfounded' into one category as these attributes contributed to such a small percentage of the overall outcome. Then, we created the donut by essentially creating two pie charts, putting one inside the other with white fill. We completed this operation because Tableau does not provide an "out-of-the-box" donut chart solution.

## Conclusion

From an analysis perspective, we were able to communicate key points to the audience describing where, when, what and how criminal incidents occurred from 2018 to 2022 within the city of San Francisco by providing effective visualizations. In part 1, We answered how incidents were reported by location, both by police district, neighborhood, and incident category. In part II, we answered when crimes are most likely to occur in the city. In part III, we described how all incidents are recorded, and how different variables effect the number of reports generated. Additionally, we summarized the results by creating a dashboard that allows the audience to consume high level familiarity with the San Francisco Police Department Incident Dataset in a short period of time while limiting cognitive overload. With a team of five members from diverse backgrounds, we were able leverage a variety of experiences and perspectives allowing us to create compelling visualizations while limiting redundancy. We shared knowledge and learnings with one-another to improve our skill set, not only with Tableau, but overall data communication skills in this course. This enabled us to deliver a valuable presentation to the audience.

**Reference:**

Jackman, T. (2020, May 20). *Amid pandemic, crime dropped in many U.S. cities, but not all*. The Washington Post. Retrieved May 15, 2022, from https://www.washingtonpost.com/crime-law/2020/05/19/amid-pandemic-crime-dropped-many-us-cities-not-all/

Ohanian, L. (2021, November 9). *Why San Francisco Is Nearly The Most Crime-Ridden City In The US*. Hoover Institution. Retrieved May 15, 2022, from https://www.hoover.org/research/why-san-francisco-nearly-most-crime-ridden-city-us

*SFPD Incident Report: 2018 to Present*. SFPD Incident Report: 2018 to Present - DataSF Dataset Explainers. (2021, July 21). Retrieved May 15, 2022, from https://datasf.gitbook.io/datasf-dataset-explainers/sfpd-incident-report-2018-to-present