Introduction:

The ability to properly communicate data analysis results using various forms of visualizations is one of the most crucial abilities for a data scientist. The goal of visualization is to make information simple to read and understand for others. As a result, too much detailed visualization should be avoided. (Metwalli, 2020). We're working with an intriguing dataset in this project because we're dealing with something that happens all the time in society. The information comes from the City and County of San Francisco. This dataset contains crime reporting events from January 1, 2018, until the present day. Every hour, the dataset is updated. According to data from the San Francisco Police Department, there has been a 32 per cent increase in car break-ins so far in 2022, with goods stolen inside. (Sierra, 2022). The dataset is updated hourly, as previously stated, however, the dataset we are using is provided by a professor from the same source. (Department, Police, 2022).

We want to obtain a better grasp of some broad crime trends as a result of this initiative. We'll also determine the most common types of offences. We'll also check into the police departments and neighborhoods in the city. In addition, we will discuss what resolutions/actions can be taken to help the police department and the general public better understand the issues.

We are utilizing Tableau as a visualization tool for our project because we are more familiar with it. Tableau and Qlik are both fantastic visualization tools, but we believe Tableau makes data exploration and analysis considerably easier.

Research Questions:

After understanding the variable and doing exploratory data analysis, we as a team had few research questions in our minds against which we created the visualizations to explore data more in other to find answers to those questions and any important insight/ trend from the dataset.

1. What is the trend of crimes reported in 2021 vs the crimes reported in 2018?
   i) What is the monthly comparison of crimes reported in 2021 vs the crimes reported in 2018?
   ii) What is the hourly trend of crimes reported in year 2021 vs those reported in year 2018?
2. What are the trends of crime occurrence frequency?
   i) What is the crime occurrence frequency in each district?
   ii) What are the Top 10 neighborhood based on incident counts?
   iii) What is the frequency of crimes' subcategories year-wise?
3. What is the quarterly crime frequency?
4. What is the quantum of crimes based on present status?

Exploratory Data Analysis:

For exploration data analysis, important parameters such as Incident Date, Incident Year, Incident Time, Incident Category, Incident Subcategory, Incident Description, Resolution, and Police District are employed extensively. We double-checked a few criteria, such as the incident date, for any missing data or formatting issues.

Following that, we looked at the association between how the data communicate inside the dataset. We discovered that some factors, such as analysis neighborhood and supervisor district, were missing data or were not yet updated. In our subsequent analysis, those characteristics did not contribute significantly, so we left them alone.

We tried to keep our visualization simple by using bar graphs, line graphs, doughnut graphs, tree diagrams, and heat maps. For our project, we've created two interactive dashboards using Tableau, which we'll present as the report progresses.

Variables:

The dataset gives a very good idea about the granularity of each incident/ record using all 34 variables. Some variables namely Incident time, date and year tells about the time of the incident reported. Whereas incident type, category and description give information about the categorization of the incident. The variables namely Police District, latitude, longitude and neighborhood gives information about the location of the incident occurred.
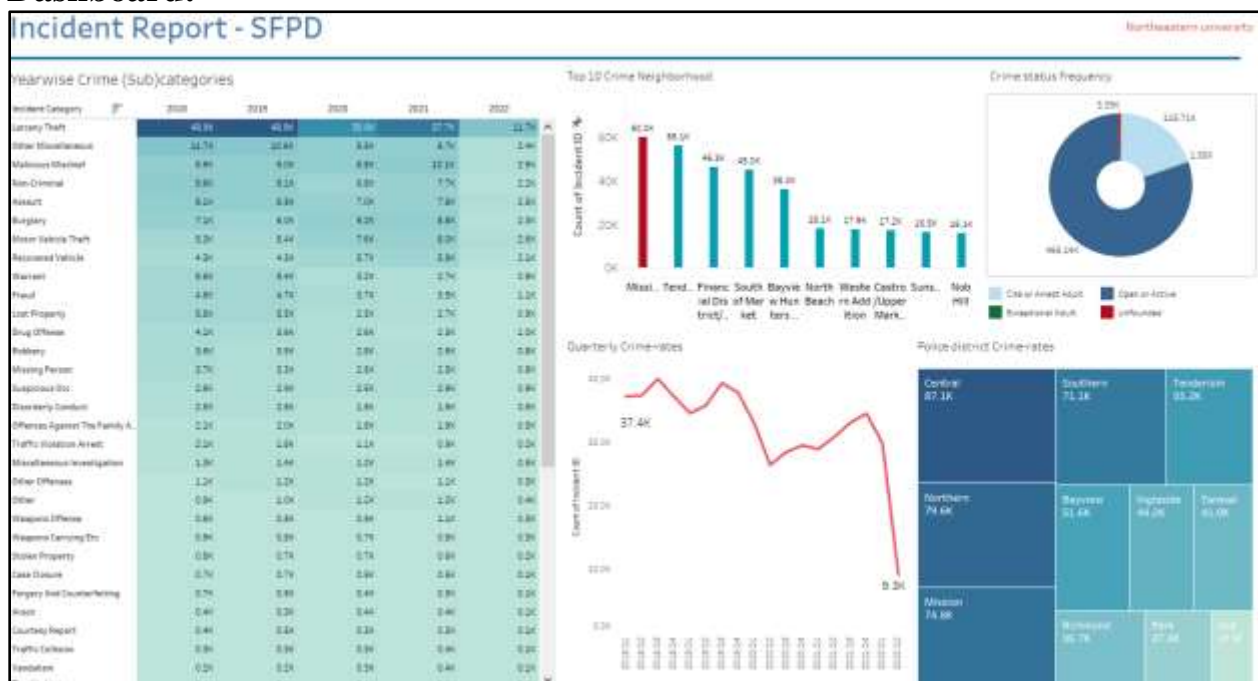
Possible Findings:

Dashboard:

Figure 1: Dashboard 1

Using tableau we've made an interactive dashboard in which we have show the trends of crimes over years available from dataset and their quarterly trends from a line graph, we get top 10 neighborhoods for the crimes giving the status of crime frequency from donut chart. And at last we created the tree map which is ranked on the frequency of crimes district wise.
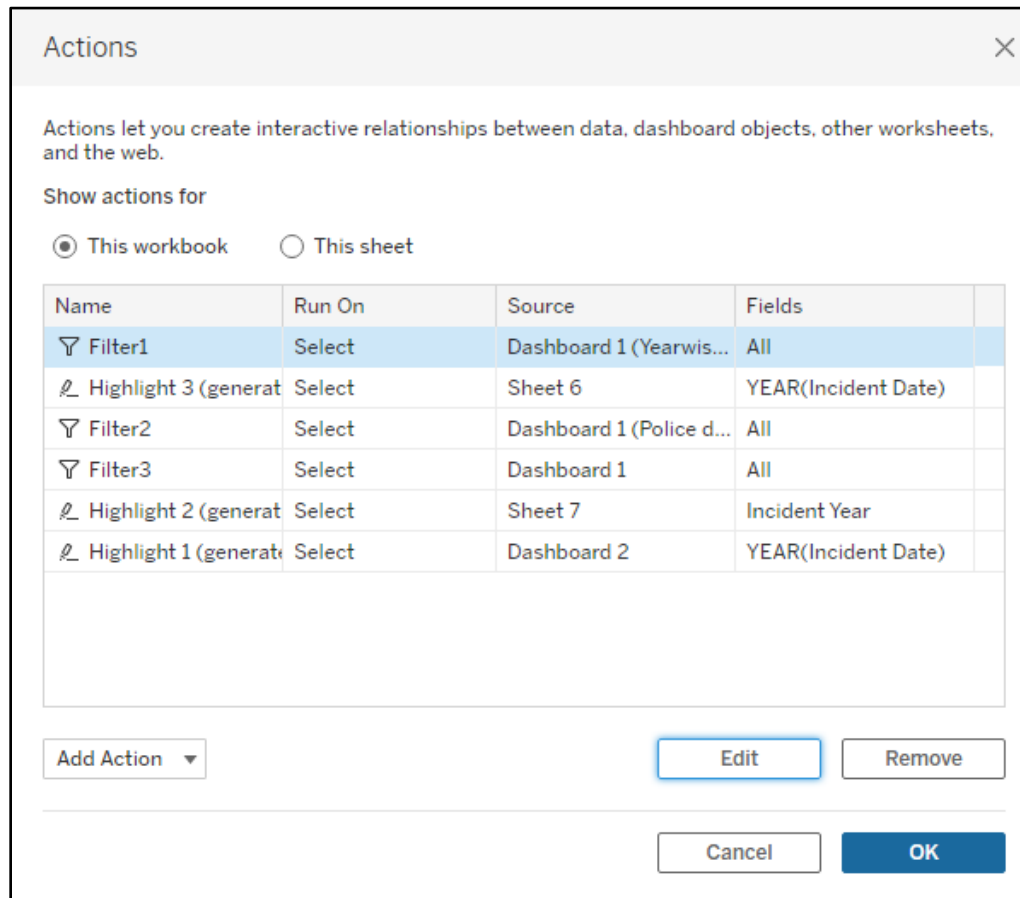
Group 1: Final Assignment Paper

Figure 2: Action window

In the table we have drilled down the sub category from the category by a click feature from 2018 till 2022. Each graph in the dashboard is linked through filters. From the filters we have actioned all the filters to run on select option rather hover to apply multiple filters and narrow down the search options.

All the trends show in general view so we have configured to show all the values once filters are cleared for the dashboard. In the process of creating the dashboard we have given a total of 3 filters

Earlier we happened to give the trends based on police district and the crime (sub)category to show all the trends. Then we moved on to give all the trends to pint point the answer to question where we checked the open/active cases/police district with the most/dangerous crimes.

Group 1: Final Assignment Paper

Figure 3: Interactive Dashboard with filters implemented

To know the trend of only open incidents for the year 2021 of mission neighborhood. We have done below steps,
1) we first clicked open cases,
2) then we clicked only mission neighborhood and
3) then clicked 2021

Here, we can see that the crimes from 1ˢᵗ quarter till 4ᵗʰ quarter in the year 2021 increased from 2.0K to 2.7K in the neighborhood of Mission.

1. What is the trend of crimes reported in 2021 vs the crimes reported in 2018?
   i) What is the monthly comparison of crimes reported in 2021 vs the crimes reported in 2018?
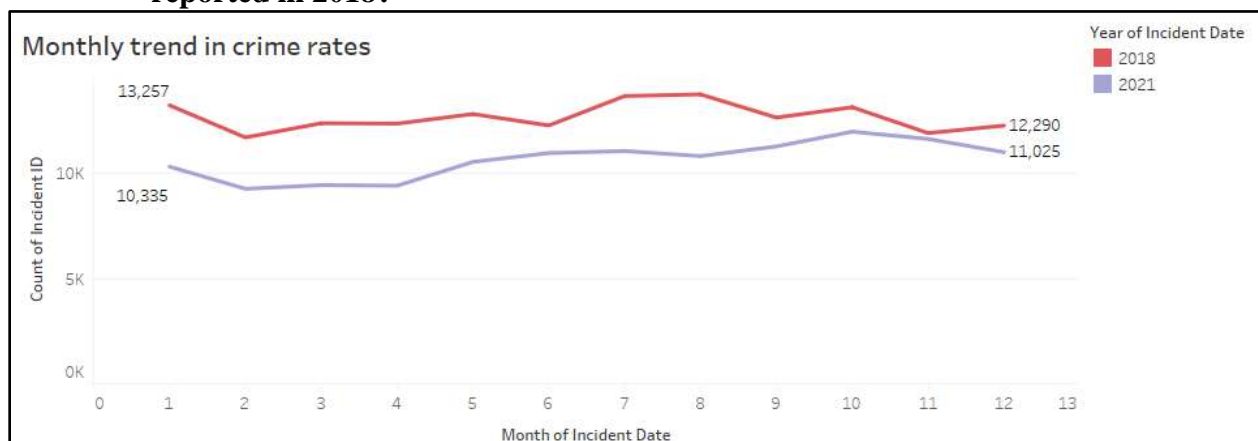

Figure 4: Monthly trend in crime rates

We selected line graph, because line graphs are best to show changes over the period of time and when we have to compare two trends then line gives us very clear idea about the differentiation.

On x-axis there are months of each year i.e. of year 2018 and the year 2021. On y-axis there is the count of the incidents.

The incidents of 2018 are shown by pink color whereas the incidents of 2021 are shown by purple color.

Here we can easily see, that overall incidents in 2021 are less as compared to the incidents reported in the year 2018. However, in the month of November the number of incidents in the year of 2018 and 2021 are almost same.

ii) What is the monthly comparison of crimes reported in 2021 vs the crimes reported in 2018?
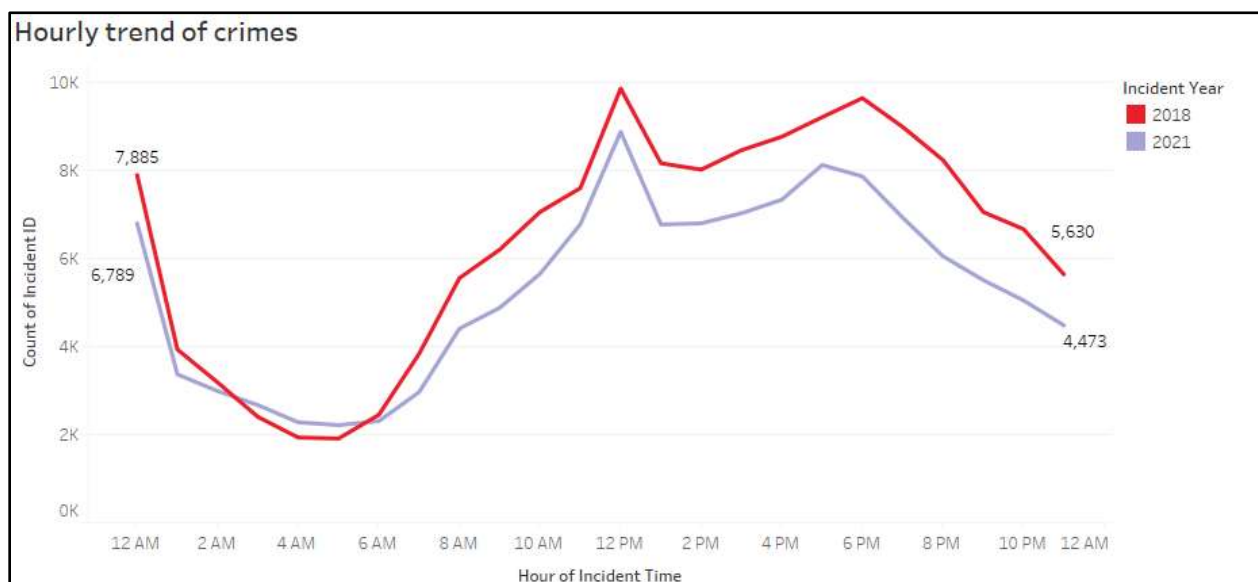

Figure 5: Hourly trend of Crimes

Since this visualization also involves a line graph and as earlier said that we selected line graph, because line graphs are best to show changes over the period of time and when we have to compare two trends then line gives us very clear idea about the differentiation.

On x-axis there are time of day (in AM/ PM format) i.e. of year 2018 and the year 2021. On y-axis there is the count of the incidents.

The incidents of 2018 are shown by red color whereas the incidents of 2021 are shown by purple color.

Here we can easily see, that overall incidents in 2021 are less as compared to the incidents reported in the year 2018. However, from 02:00 AM till 06:00 AM the number of incidents in the year 2021 have been higher as compared to that reported in the year 2018.

Group 1: Final Assignment Paper

However, there is a general trend around 12 PM, that show peak in the number of crimes.

2. What are the trends of crime occurrence frequency?
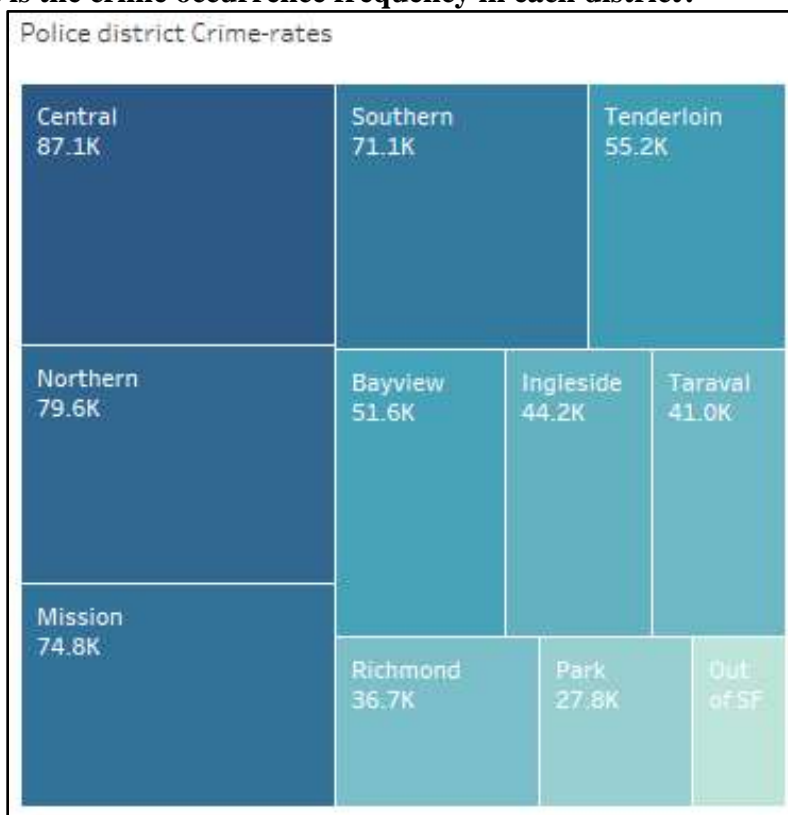   **i) What is the crime occurrence frequency in each district?**

Figure 6: Police District Crime-Rates

We selected tree map, because the size and proportion of the data is best represented using tree map. It is also good to use tree map when data is well structured. Tree map makes it easy to understand the data presented, you can easily know the largest area covered by any attribute is the highest among all other attributes.

The depth/ darkness of the blue shade and the size of the box depicts the quantum of crimes in those particular police district. Here we can see that the greatest number of crimes are reported in the Central district of San Francisco. Whereas the least number of crimes are out of San Francisco.

Group 1: Final Assignment Paper

ii) What are the Top 10 neighborhood based on incident counts?
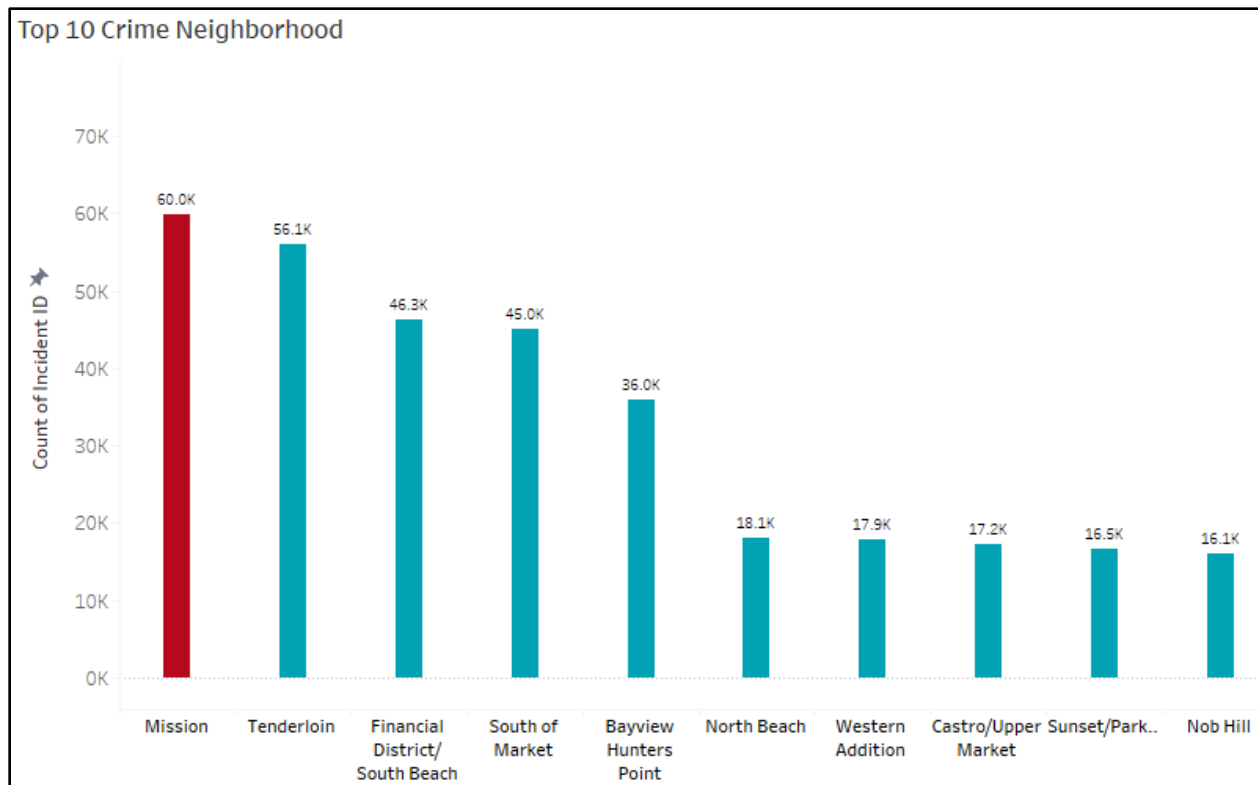
Figure 7: Top 10 Crime Neighborhood

We have selected bar graphs because bar graphs are best when we have to compare stats between different groups. Here we can see the quantum of incidents in the respective neighborhoods. On x-axis there are neighborhoods whereas on y-axis there is the count of the incidents. The above bar graph shows only the top ten neighborhoods, where the incidents have most occurred.

The bar of the neighborhood with most incidents is shown in red color whereas all other neighborhoods (or second to tenth in ranking) are shown in blue color. Based on the neighborhood analysis, it would be easier for the Police Station Division to analyze which areas to concentrate more on. From the analysis, we can see that the Mission neighborhood has the highest number of incidents reported and followed by Tenderloin, Financial District, South of Market, and Bayview Hunters Point being the top 5 neighborhoods with the highest incidents reported. For this Analysis, we have taken neighborhoods on the x-axis and the number of incidents reported on the y-axis.

iii) What is the frequency of crimes' subcategories year-wise?

Incident Category	2018	2019	2020	2021	2022
Larceny Theft	48.9K	48.9K	30.6K	37.7K	11.7K
Other Miscellaneous	11.7K	10.6K	8.5K	8.7K	2.4K
Malicious Mischief	8.9K	9.0K	8.8K	10.1K	2.9K
Non-Criminal	9.6K	9.1K	6.8K	7.7K	2.2K
Assault	9.1K	8.9K	7.0K	7.8K	2.5K
Burglary	7.1K	6.0K	9.2K	8.8K	2.3K
Motor Vehicle Theft	5.3K	5.4K	7.6K	8.0K	2.6K
Recovered Vehicle	4.3K	4.3K	5.7K	5.9K	2.1K
Warrant	5.6K	5.4K	3.2K	2.7K	0.9K
Fraud	4.8K	4.7K	3.7K	3.5K	1.1K
Lost Property	5.8K	5.5K	2.5K	2.7K	0.9K
Drug Offense	4.1K	3.6K	2.6K	2.3K	1.0K
Robbery	3.6K	3.5K	2.8K	2.6K	0.8K
Missing Person	3.7K	3.3K	2.6K	2.5K	0.8K
Suspicious Occ	2.8K	2.9K	2.5K	2.9K	0.9K
Disorderly Conduct	2.5K	2.8K	1.8K	1.9K	0.6K
Offences Against The Family A...	2.1K	2.0K	1.8K	1.9K	0.5K
Traffic Violation Arrest	2.1K	1.8K	1.1K	0.9K	0.2K
Miscellaneous Investigation	1.3K	1.4K	1.2K	1.4K	0.5K
Other Offenses	1.1K	1.2K	1.2K	1.1K	0.3K
Other	0.9K	1.0K	1.2K	1.2K	0.4K
Weapons Offense	0.8K	0.8K	0.9K	1.1K	0.3K
Weapons Carrying Etc	0.9K	0.8K	0.7K	0.8K	0.3K
Stolen Property	0.8K	0.7K	0.7K	0.6K	0.2K
Case Closure	0.7K	0.7K	0.5K	0.5K	0.1K
Forgery And Counterfeiting	0.7K	0.6K	0.4K	0.3K	0.1K
Arson	0.4K	0.3K	0.4K	0.4K	0.1K
Courtesy Report	0.4K	0.5K	0.3K	0.3K	0.1K
Traffic Collision	0.3K	0.3K	0.3K	0.4K	0.1K

Figure 8: Year-wise Crime (Sub)categories

We required numbers and graphics for the incident subcategory, therefore we chose a table with a heat map function. Because it's part of an interactive dashboard, it changes depending on the variables you choose. For example, we can select a district and then verify a specific neighborhood within that district. All essential descriptions will be presented in detail, along with an update on a doughnut chart, and the heat map will be updated as well. Several instances were reported to the San Francisco Police Department, as seen in the diagram below. We also wanted to look at the trend of occurrences reported over time for a certain category and sub-category with this visual.

It's a simple table with Incidents columns and incident sub-columns on the x-axis and year on the y axis with the number of incidents recorded. With 28.5K incidents recorded in 2018, larceny from cars was the most common sub-category. The category of larceny-theft has the highest number of records. The heat map is represented by the legend. The hue darkens as the number of incidents rises, and vice versa.

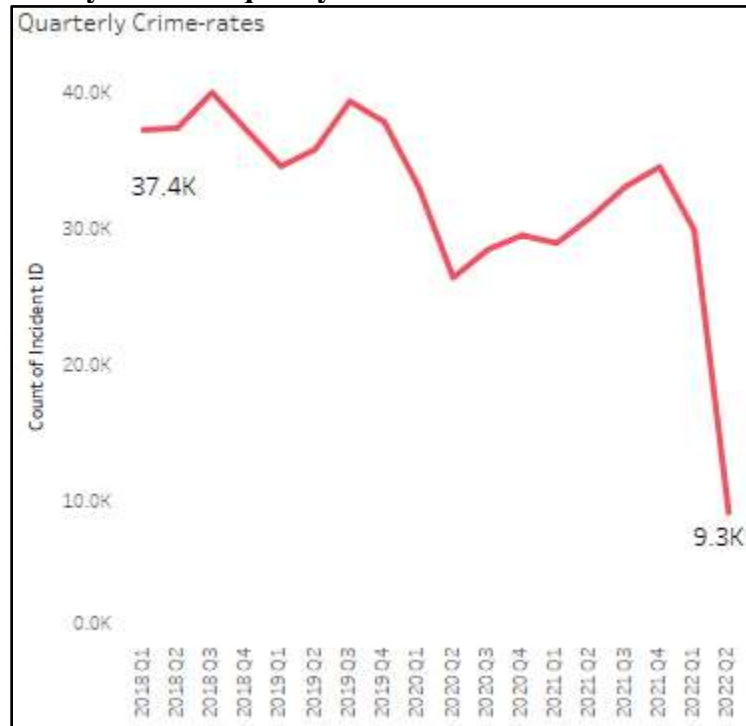### 3. What is the quarterly crime frequency?

Figure 9: Quarterly Crime-rates

We have selected line graph to depict against the research question, because it is easier to know the trend and at a glance, we can know that the trend is downward. On x-axis there are yearly quarters from the year 2018 till the year 2022. The above bar graph shows the quarterly trend of the crimes/ incident reported. Based on the Quarterly Crime Rate analysis, it would be easier for the Police Station Division to analyze in which quarter crime rates are high, Also, with this visual, we wanted to look at the trend of the incidents reported over years for a given category, it is a simple table representation with Incidents columns and incident on the x-axis and year with the number of incidents reported on the y axis. 2018 is the highest number of crimes are reported, in this Q3 is high after that 2019Q3 is crimes too high, After that police division is work too hard solve the cases 2020Q2 cases low and pregnant cases too low compared with the previous year.

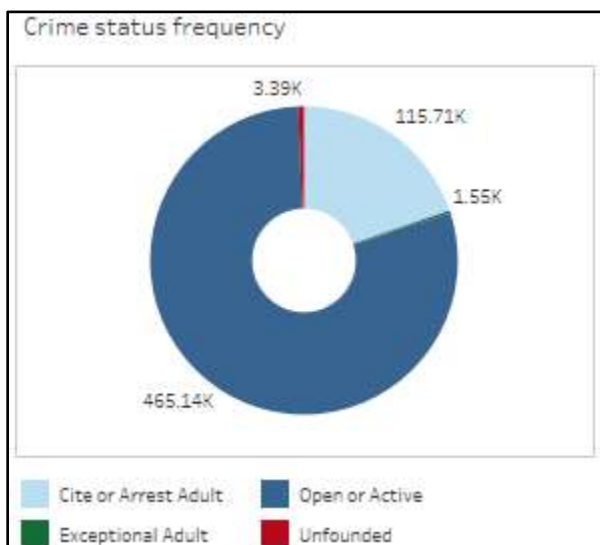**4. What is the quantum of crimes based on present status?**

Figure 10: Crime Status Frequency

As we can see that we have several incidents reported to SFPD, it is essential to look at the Crime Status Frequency categories in which the incidents are reported. Also, with this visual, we wanted to look at the trend of the incidents reported over years for a given category Cite or Arrest Adult, Exceptional Adult, Open or Active, Unfounded. It is a simple Pie Chart representation Total Open or Active Cases are 465.14K Most of the crimes are doing Adults most of the cases are reported as larceny-Vehicle and larceny-theft, The Larceny-Vehicle category holds the top position for the highest number of incidents reported.

Design Decisions:

For first question, to show the crime trend over the period of time, we used line graphs, because line graphs are the best for displaying changes over time, especially for comparing two trends, they provide a clear picture of the differences.

For first part of second question, the tree map was chosen to show the incident quantum in each district. We used tree map because it best represents the magnitude and proportion of the data. When data is well structured, tree maps are also useful. Tree maps make it simple to comprehend the information displayed; you can quickly see which attribute has the most area covered among the others.

For second part of second question, to show incident subcategories, it involved lots of numbers against each year, therefore we used a table with a heat map function. It changes colors depending on the number been shown in comparison with other values of same attribute.

For third question, we chose a line graph to represent the quarterly crime frequency because it is easy to see the trend and we can see that the trend is decreasing at a glance.

For the fourth question, we have used a donut chart to represent open/ active incidents, unfound and others because donut charts are good at representing proportions of data, its to understanding while looking at the size of the portion shown.

Conclusion:

The study above demonstrates how effective visualizations are in conveying information and answering difficult problems. The pictorial representation not only answers the questions but also makes it appealing to the reader since images are more successful at delivering information than numbers alone.

We attempted to explain difficult query questions using simplified visuals in this project using Tableau software. We used line charts, a tree map, a bar graph, and a heat map table to depict some year-by-year trends. A line graph and a doughnut chart were used to communicate the sub categorical question. The dashboard we created is interactive, and the project "twbx" file is included in this report so that you may view and interact with the data. According to our report's monthly crime statistics, the crime rate in 2021 is significantly lower than in 2018. Summertime appears to have higher crime rates than other times of the year. We also discovered that there is a high incidence of crime from 4 a.m. to 6 p.m., with a peak at midday, by looking at the hourly pattern of crimes. The Central district of San Francisco is the most crime-prone, followed by the Northern, Mission, and Southern Districts. The above-mentioned districts account for nearly half of all recorded crime events. Whereas the Mission neighborhood in San Francisco's Mission district tops the list for the most criminal incidences reported.

Larceny theft is the most commonly committed crime. In the area of larceny-theft, there is a subcategory of car break-ins and theft that is recorded the most, with the category topping the list with a startling 48.9K reported events in 2018, which we can see from the table and heat map is steadily decreasing. By 2022, the number of recorded events has dropped to 11.7K. Overall, we can see that the number of offenses reported has decreased nearly fourfold from the year 2018 to the present. As discussed earlier we have created an interactive dashboard. This dashboard can be valuable for police departments because it displays all statistics at a glance. Because it is interactive, users can select any district, area, or kind of crime to see the most recent trends. A report's status, whether open, active, closed, or unfounded, can also be checked.

Reference:

Metwalli, Sara A. "Data Visualization 101: 7 Steps for Effective Visualizations." Medium, Towards Data Science, 8 Sept. 2020, https://towardsdatascience.com/data-visualization-101-7-steps-for-effective-visualizations-491a17d974de.

Department, Police. "Police Department Incident Reports: 2018 to Present: DataSF: City and County of San Francisco." Police Department Incident Reports: 2018 to Present | DataSF | City and County of San Francisco, 4 May 2022, https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783

Knaflic, Cole N. (2015). Storytelling with Data: A Data Visualization Guide for Business Professionals. Hoboken, NJ. John
Wiley & Sons. ISBN: 978-1-119-00225-3

Sierra, Stephanie. "'Honest Answer': Here's Why One of SF's Top Officers Says Crime Is Here to Stay." ABC7 San Francisco, KGO-TV, 9 Apr. 2022, https://abc7news.com/san-francisco-crime-sfpd-response-time-sf-car-break-ins/11727958/

Appendix:

Along with this report we have submitted the Tableau (.twbx) file namely,
Group1_Final_Assignment.ALY6070.twbx.