Final Project

ALY 6070 – Communication and Visualization for Data Analytics

Santhi Kiran Chavali

Professor Name: Jack Bergersen

Northeastern University, Boston

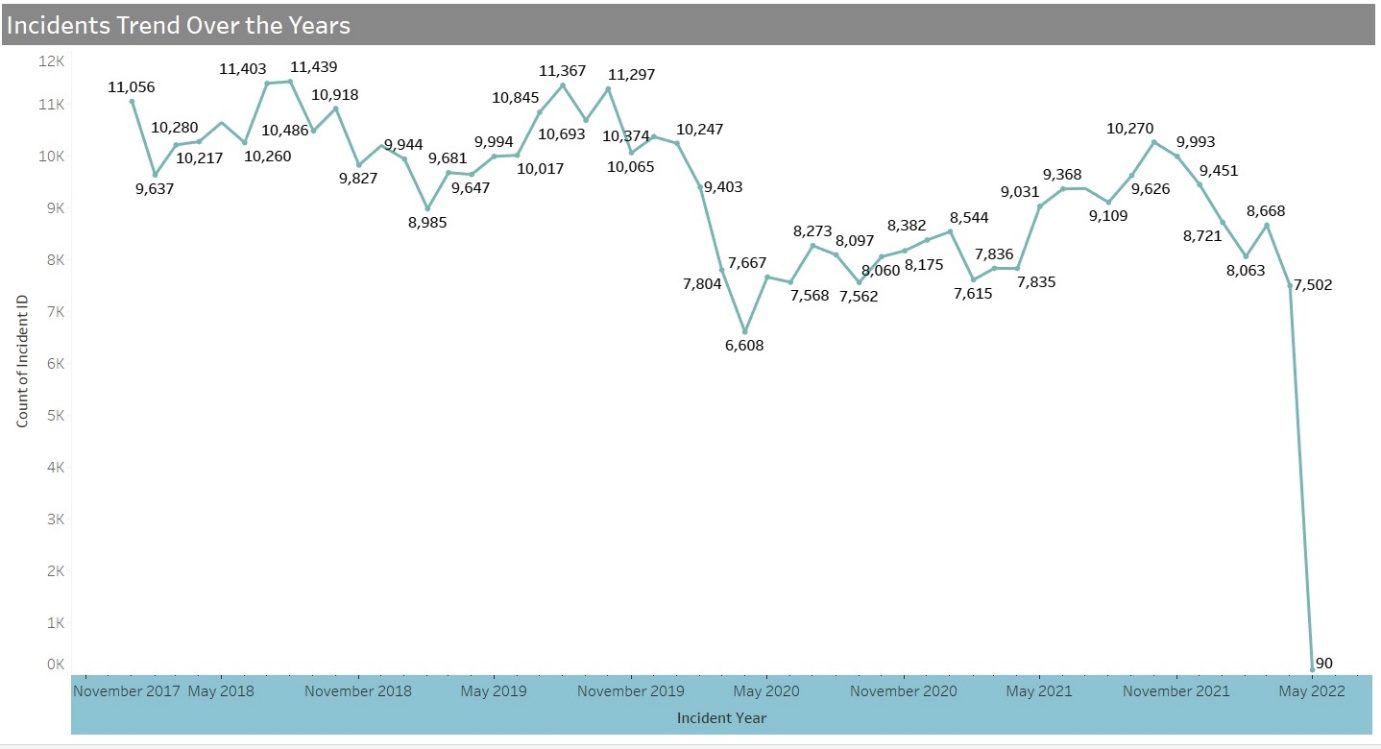


**Introduction**

For this final project assignment, we will be using The San Francisco Police Department incident report dataset provided by the instructor. The dataset for incident reports does a decent job of giving counts of incident reports broken down by category, date, time, and location. The amount, location, and type of criminal incidents reported to or reported by the SFPD can all be better understood with the aid of this data. To make incident reports across neighborhoods, police districts, and supervisory districts more understandable, geographic information is provided in an anonymous manner. The study of data trends over time is made possible by the date information provided. We have used Tableau for this assignment to visually represent the data

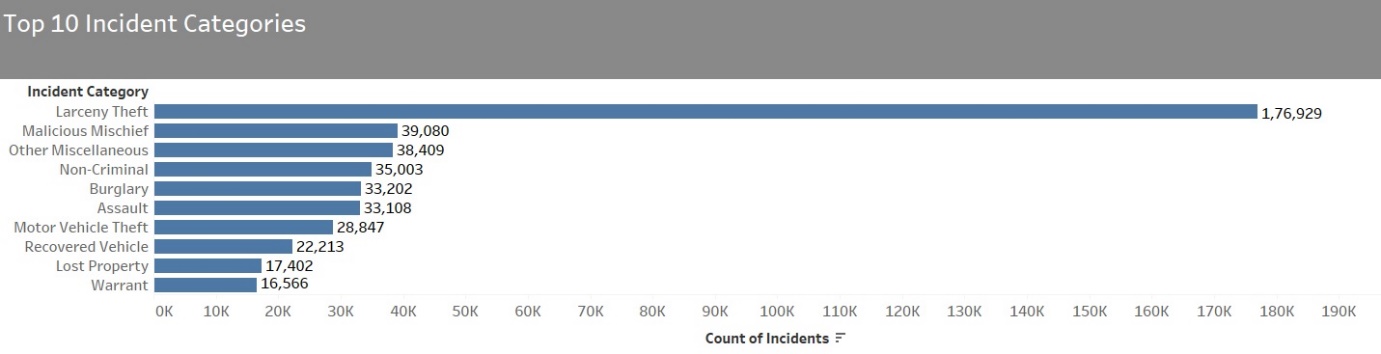
**Research Questions and Answers**

1. **What has been the trend of the incidents over the years?**



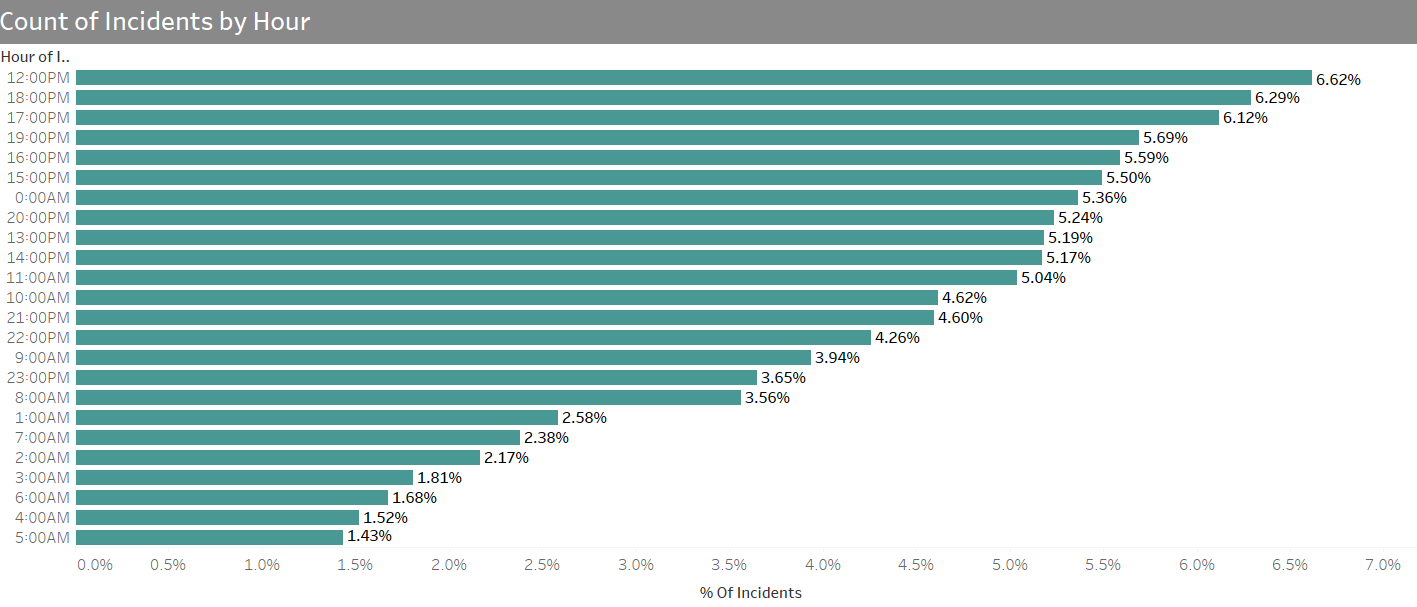
To answer this question, we have used the line graph which is useful to represent the trend followed in the dataset. On the x-axis, we have Month and Year and, on the y-axis, we have the count of incidents in thousand. Even though the goal is to show the trend of the incidents, just in case if the audience is interested to know the count. we have labeled the count on the line graph as well. From the above graph, we can observe that from November 2017 to November 2019 on average the count of crimes remain the same. But, after November 2019 we can see that there is a fall in crimes still in 2020 and followed the pattern till November 2021. Crimes drastically dropped later might be due to covid pandemic. When considering the overall picture even excluding the drastic drop in the end we can see that the total count of crimes is decreasing.

1. **What are the top ten types of incidents that occur in districts?**



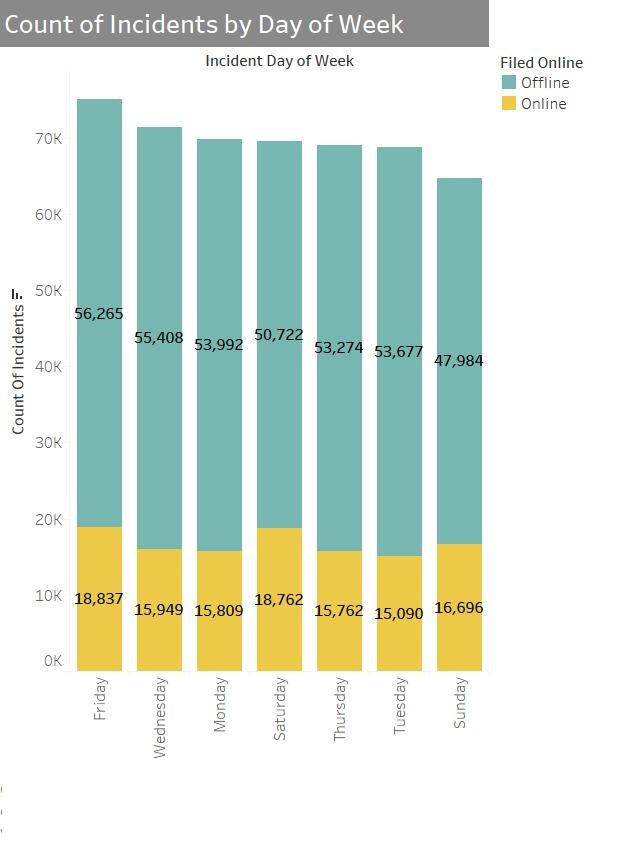
We have chosen the horizontal bar graph to answer this question. The bar graph is simple and easy to understand and it can be used to summarize a large set of data in a simple visual form. On the x-axis, we have the count of incidents in thousands, and on the y-axis, we have the category of the incident (top 10). Each bar graph is labeled with a count at the end of the bars. From the above bar graph, we can say that Larceny theft is the most often occurred crime and it is in the first place in the top 10 list with a huge number of 1,76,929 incidents followed by Malicious Mischief with 39080 and warrant being the last in the top 10 list with 16566 incidents. This graph allows the audience to focus on what type of crimes to be addressed to manage the crime rate. ***We have added action filter for the top 10 incident categories which will filter the resolution trend graph based on the selection.***

1. **At what time of the day are the incidents occurring more?**



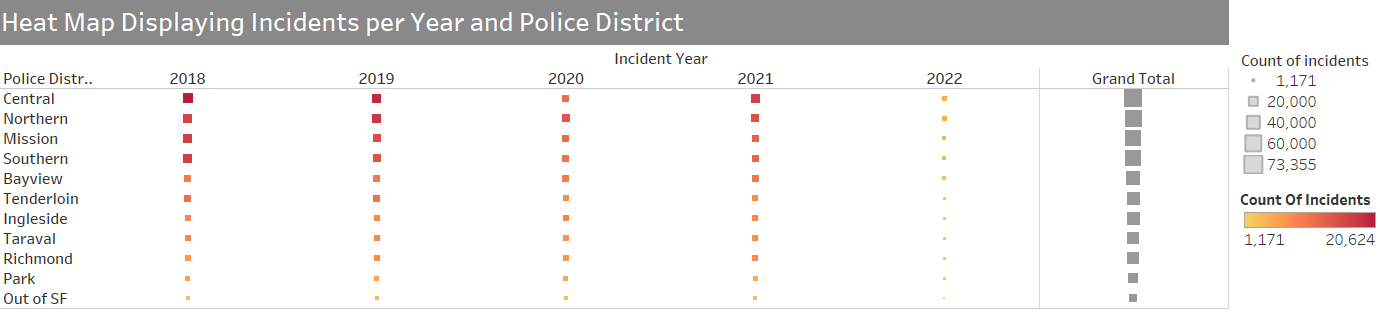
This question is framed to know at what time of the day the incidents are taking place. To answer this question we have used the horizontal Bar graph which represents the data perfectly and displays how much percentage of total incidents are taking place at each hour. The X-axis indicates the percentage and the Y-axis indicates the hour of the incident. Each bar is labeled with a percentage at the end of the bar graph and bars are sorted in descending order. From the above horizontal bar graph, it is observed that most incidents take place at 12:00 PM with 6.62% followed by 18:00PM with 6.29%, and 5:00 AM being the last with incidents occurring at 1.43% which shows less chance to occur crime. This graph allows the audience to make a decision on at what time of the day the security has to be increased. ***We have used use as a filter property to apply filters to the remaining reports on the dashboard.***

1. **Which days are riskier for crime to happen?**



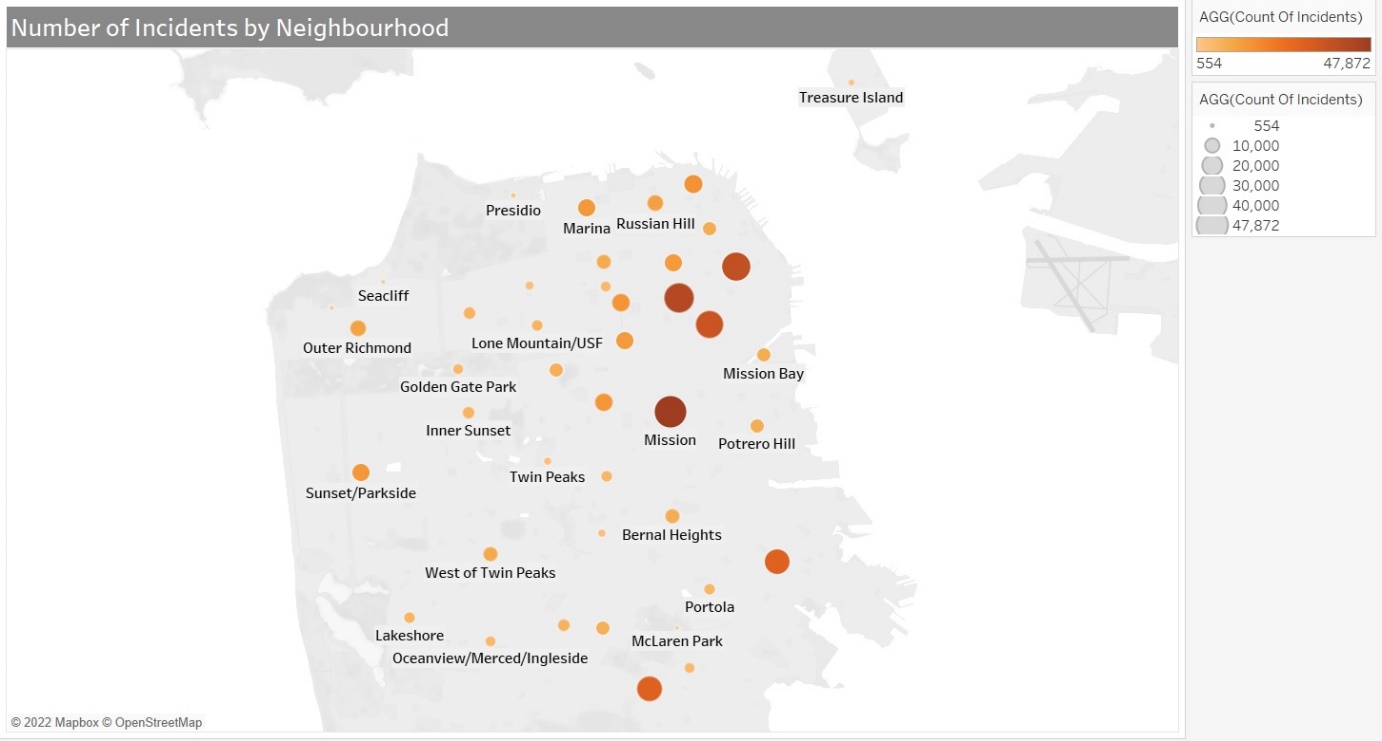
This question is asked to know if is there any particular day when the crimes occur more. To answer this question we have used a stacked vertical bar graph instead of a standard bar graph because basically, we have two types of complaints received online and offline, So, using a stacked bar graph is easy to represent this type of data. On the x-axis, we have the days of the week and on the y-axis, we have the count of the incidents in thousands. We have used two different colors in each bar to represent online and offline complaints and the same is also shown using the legend for better understanding. Each bar is labeled inside with a count of the incidents. From the above graph, considering the number of incidents that occurred, it is observed that there is no major difference between the days of the week. However, when compared to other days, crimes are taking place a little more on Fridays and Sundays are the last.

1. **What is the volume of incidents per year by police districts?**



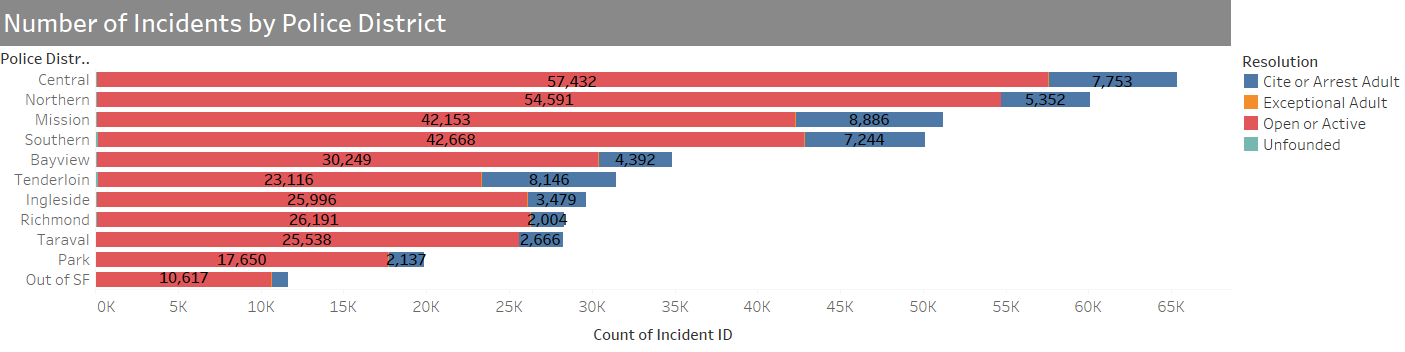
To answer this question a heat map is selected. Heat Maps help viewers focus on the parts of data visualizations that matter most by helping to better visualize the volume of locations and events inside a dataset. We have used one color for the whole graph and with different shades and the same is displayed using the legend. The Size and thickness of color in the squares in the above graph display the count of incidents each year by police districts. More thickness of the color more number of incidents in that particular year in the respective police districts/ From the above heat map we can observe that the count decreased in all the police districts in the year 2022 and over the 5 years, Crimes in the Central police district is more except in the year 2020 where more number of cases are registered in Northern Police District.

1. **What is the number of incidents occurred in Neighbourhood?**



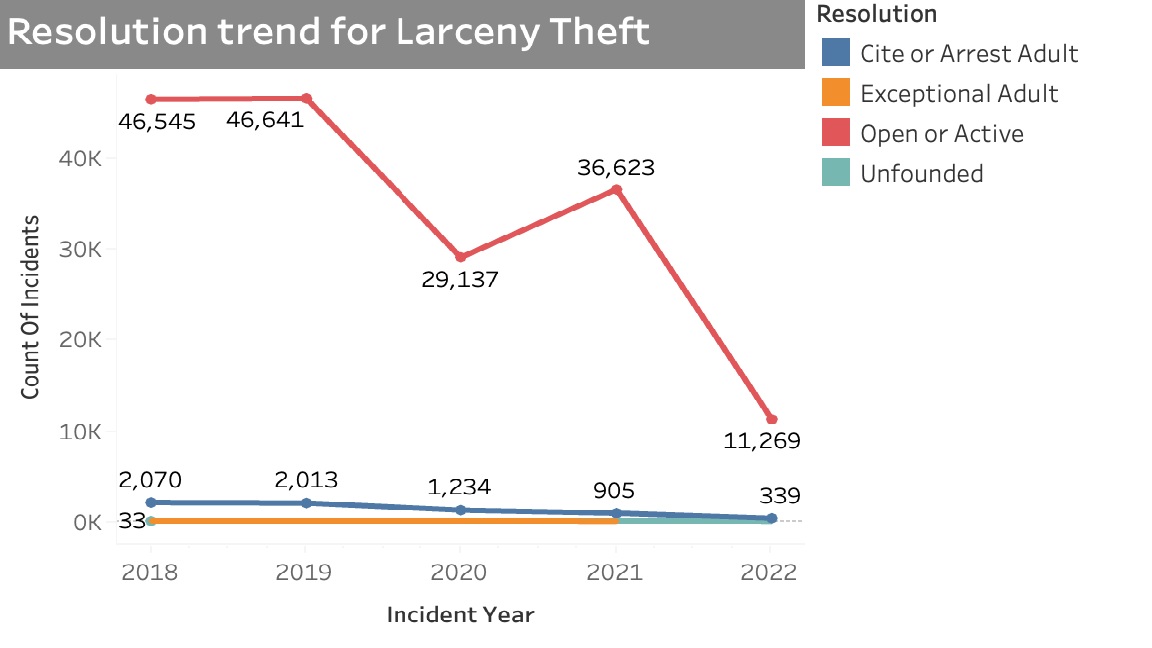
The scope of this question is to identify the neighborhood where each of the incidents has taken place. To answer this question, we have used a map graph as they are very useful when we need to thoroughly depict data and compare it across regions for strategic decision-making. Map charts offer crucial setting options that assist users in seeing local data and transforming it into informative data. Maps are a useful visual aid for explaining scale, position, and distance to the viewer, and also we believe that a map chart gives an overall idea for the audience to know in which location most of the crimes were committed. Neighborhoods are identified using the longitudes and latitudes and only one color with different shades is used to show the count of incidents. The thicker the color and large the size of the points the more the incidents occurred. The same is displayed using the legend. From the above map chart, we can observe that the Mission neighbourhood are having the more incidents and Lincoln park neighbourhood has the least

1. **What is the number of incidents by police district and what is the resolution status?**



To answer this question we have used the horizontal stacked bar chart which is used to represent multiple variables in a single bar for different categories. The X-axis indicates the count of incident IDs and the Y-axis indicates the Police districts. Each stacked bar consists of the number of crimes in each police district and it also includes the status of the resolution as well. We have used different colors for the resolution status that is also displayed using a legend and each bar is labeled with the counts of resolution and sorted using counts in descending order. From the above bar graph, it is observed that the Central police district is having the maximum number of crimes and the Out of SF being the least. It is also observed that in all the police districts the maximum number of cases are in the open or active status followed by cases in cite or arrest audit status and in all the police districts, Exceptional audit and Unfounded cases are in negligible counts.

1. **What is the resolution status of the most popular crime committed – Larceny theft, over the years?**



We have used a line graph with multiple lines to represent the data which is a perfect option to visually display the trend followed. On the X-axis we have Incident Years and on the y-axis we have a count of the incidents. Four different colored lines are used to represent the status of the resolution. We have also labeled the line with the exact count of incidents. From the above graph, we can see that cases in open and active status are still more over the years. However, we can observe that there is a decrease in the count by 2022. Count of Cite or arrest adult status is gradually decreasing over the years. Default view for the resolution trend graph for larceny theft, based on the action filter selection from the top 10 incident categories, filter will be applied accordingly

**Conclusion**

For this final project assignment, we used Tableau to conduct an analysis on the San Francisco Department Incident report. Various types of graphs are used to visually represent the analyzed data like line graphs, horizontal, vertical, and stacked bar charts, Heap maps and Geographical Map chart. From the above analysis we can conclude the following

* Larceny theft is the most likely of the top ten categories of incidents. The Central region has seen the most thefts with 37092 thefts, followed by the Northern region with 32352 thefts and the Southern region with 20374 thefts.
* From 2017 to 2022, there has been a significant drop in crime rates in the years 2019 to 2020, and a critical drop from November 2021 to May 2022.
* The Central, Northern, and Mission districts are the most likely to have incidents, followed by the Southern district.
* Most of the crimes are taking place at 12:00 PM and Friday is having the maximum number of crimes.

References:

1. SFPD Incident Report: 2018 to Present - DataSF Dataset Explainers. (2018). Gitbook.io. <https://datasf.gitbook.io/datasf-dataset-explainers/sfpd-incident-report-2018-to-present>
2. San Francisco Incident Reports (2018-present). (n.d.). Www.kaggle.com. Retrieved December 11, 2022, from <https://www.kaggle.com/datasets/vivovinco/san-francisco-incident-reports-2018present>
3. NetSuite.com. (n.d.). Heat Maps: Types & Benefits. Oracle NetSuite. <https://www.netsuite.com/portal/resource/articles/erp/heat-map.shtml>
4. 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