### Date: 18-05-2023

**ALY6070 - Communication and Visualization for Data Analytics Final Assignment Paper**



**NORTHEASTERN UNIVERSITY**

**College of Professional Studies, Northeastern University, Boston, MA, 02215.**

## Submitted By GROUP 3

### Satya Viswa Pavan Ranga

**Yuva Veera Prasanna Lakshmi Kocherla Rachana Maguluri**

**Yerra Sai Pavan Devulapalli Pavan Teja**

**To**

**Prof Dr Jack Bergersen**

**Final Assignment Paper**

1. | P a g e

|  |  |
| --- | --- |
| **Contents** | **Page No.** |
| **1. Introduction** | **03** |
| **2. Analysis** | **04** |
| **2.1. Research Questions** | **04** |
| **2.2. Methods used** | **06** |
| **2.3. Key Findings** | **07** |
| **3. Conclusion** | **12** |
| **4. References** | **14** |

**Final Assignment Paper**

1. | P a g e

# Introduction

The San Francisco Police Dataset is a comprehensive collection of incident reports recorded by the San Francisco Police Department. This dataset comprises 554,643 records from the years 2018 to 2022 and provides valuable insights into the patterns, trends, and characteristics of incidents in the city.

The dataset contains various attributes that describe each incident. These attributes include the date and time of the incident, the year in which the incident occurred, the day of the week, and the type of report filed. Additional information is available regarding the incident category, resolution status, police district, supervisor district, geographical coordinates (latitude and longitude), and time category (e.g., morning, afternoon, night). The dataset offers a comprehensive overview of the incidents, enabling in-depth analysis and exploration.

The Incident Date and Incident Time fields provide information on when each incident took place. The Incident Year variable allows for an examination of the distribution of incidents across different years within the dataset. The Incident Day of Week attribute enables analysis of incident patterns based on specific days. The Report Datetime field specifies the date and time when the incident was reported to the police, allowing for tracking of reporting trends.

The dataset includes attributes related to the type of report filed. The Report Type Code and Report Type Description variables provide information on whether the report was an initial report or a supplemental report. The Filed Online attribute indicates whether the report was filed online or through another channel.

The Incident Category field classifies incidents into different categories, such as larceny theft, motor vehicle theft, warrant, and more. The Resolution variable indicates the status of incident resolution, such as whether the incident is open or active or if an adult was cited or arrested.

Information about the police district and supervisor district associated with each incident is captured in the Police District and Supervisor District attributes. These variables allow for an exploration of incident distribution across different areas of San Francisco. The dataset also includes geographical coordinates (latitude and longitude) for each incident, enabling spatial analysis and mapping.

Additionally, the dataset provides information on the current supervisor districts and police districts associated with each incident. These variables offer insights into the jurisdictional boundaries and administrative divisions within the city.

By leveraging the rich attributes within the San Francisco Police Dataset, researchers, analysts, and policymakers can gain a deeper understanding of crime patterns, devise effective strategies for crime prevention and law enforcement, and contribute to enhancing public safety in the city. This report aims to explore and analyze the dataset to extract meaningful insights that can address crime-related challenges and improve community well- being.

**Final Assignment Paper**

1. | P a g e

# Analysis

## Research Questions:

### What are the top six crimes in San Francisco, and what patterns can be identified in their occurrence and characteristics?

The analysis of the graph depicting the top six crimes in San Francisco reveals interesting patterns and distribution of these crimes. Assault, Burglary, Larceny Theft, Malicious Mischief, Non-Criminal, and Other Miscellaneous are identified as the prominent crime categories. Fridays consistently had the highest number of incidents across all crime categories, with Larceny Theft being the most prevalent crime. The visualization of incident counts provides a clear representation of the relative frequencies of these crimes. These insights contribute to a better understanding of crime patterns, enabling law enforcement agencies and policymakers to devise targeted strategies for crime prevention and resource allocation.

Variables used:

* + Incident Category
  + Incident.Day.of.Week
  + Incidents Count

### How have the categories of crime in San Francisco changed from 2018 to 2022, and what are the underlying factors contributing to these changes?

The analysis of crime categories in San Francisco from 2018 to 2022 reveals notable trends. The category labeled "Others" had the highest incident count but decreased significantly.

Larceny Theft remained relatively stable. Malicious Mischief, Assault, Non-Criminal, and Burglary showed fluctuations, while Motor Vehicle Theft, Warrant, Fraud, and Recovered Vehicle decreased. These changes may be influenced by factors like law enforcement strategies, socioeconomic conditions, and community initiatives. Further investigation is needed to identify specific contributing factors.

Variables used:

* + Incident.Date
  + Incident.Category
  + Incident.Count

### How has the overall crime rate in San Francisco evolved from 2018 to 2022, and what factors have influenced these trends over the specified time period?

The analysis of the lollipop chart depicting the crime rate in San Francisco from 2018 to 2022 reveals a positive trend in the reduction of overall crime. The crime rate gradually decreased from 144,281 incidents in 2018 to 36,940 incidents in 2022. This decline suggests successful

**Final Assignment Paper**

1. | P a g e

efforts in addressing and mitigating crime in the city. Factors such as changes in law enforcement strategies, community initiatives, economic conditions, and social factors may have influenced these trends. The visually informative chart illustrates the magnitude of the crime rate each year. These findings provide valuable insights for policymakers and law enforcement agencies to develop targeted strategies and interventions to further reduce crime in San Francisco.

Variables used:

* + Year
  + Crime Rate
  + Incident Category

### What factors contribute to variations in the distribution of registered cases across different districts of San Francisco?

The stacked bar graph depicting the distribution of registered cases across different districts of San Francisco provides insights into the factors influencing the variations observed in case distribution. The graph highlights that the Tenderloin district has the highest number of cases registered in the Morning, Afternoon, and Night time periods, followed by the Mission, Central, and Southern districts. This suggests a potential correlation between district and timing of reported cases. The visualization allows for easy comparison of case counts, aiding policymakers and law enforcement agencies in developing targeted interventions. Further analysis is needed to explore underlying factors such as socioeconomic conditions and policing strategies influencing the distribution of cases.

Variables used:

* + Time\_Category
  + Police.District
  + Case Counts

### How does the number of reported cases vary in relation to changes in police incidents over time in San Francisco?

The graph analysis reveals variations in the number of reported cases in relation to changes in police incidents over time in San Francisco. Fluctuations in reported cases across different police districts, including Bayview, Central, Ingleside, Mission, Northern, Out of SF, Park, Richmond, Southern, Taraval, and Tenderloin, indicate variations in incident occurrence and reporting in different areas. While some periods show a correlation between reported cases and police incidents, further investigation is needed to understand the underlying factors. The graph visually compares the magnitudes of incidents using bars and a scale ranging from 20,000 to 140,000. This information provides insights for effective law enforcement and crime prevention strategies.

**Final Assignment Paper**

1. | P a g e

Variables used:

* + Police District
  + Incident.Date
  + No of Incidents- Count(Incident.Category)
  + Report.Datetime

## Methods Used

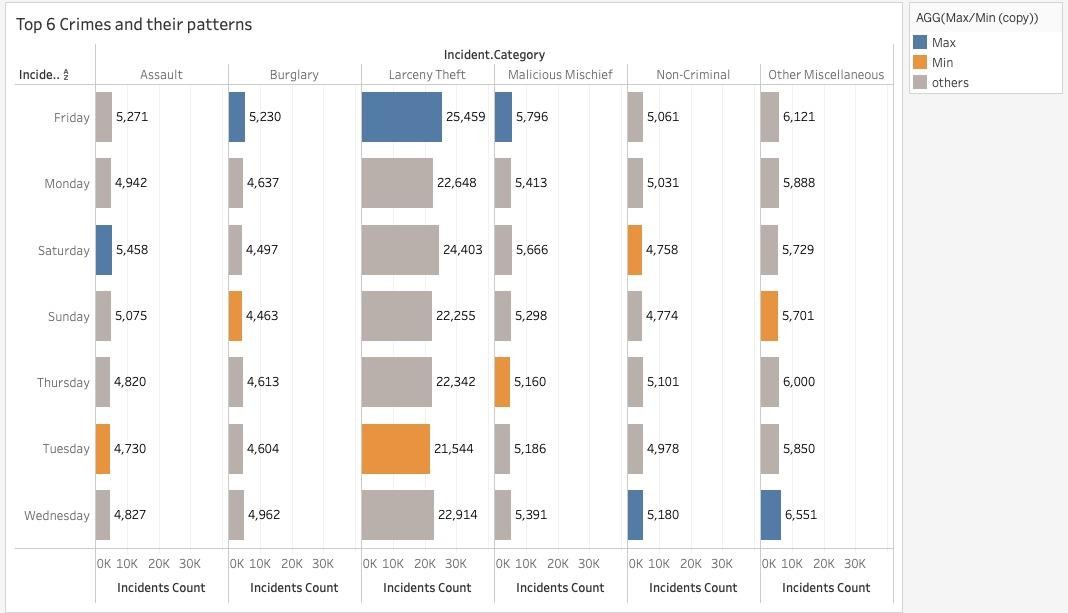
* + 1. **Grouping and aggregation:** Data was grouped and aggregated based on relevant variables to derive meaningful insights. Grouping variables, such as incident category, incident day of the week, year, police district, and time category, were utilized to organize and analyze the data effectively.
    2. **Filtering and data segmentation:** Filters were applied to focus on specific time periods, geographical regions, or other relevant subsets of the data. Segmentation of data based on different factors enabled the exploration of patterns, trends, and variations within specific contexts.
    3. **Calculation and computation:** Calculated fields and computed measures were used to derive additional insights or perform complex calculations. Examples include calculating crime rates, determining relative frequencies, and analyzing changes over time.
    4. **Customization of visualizations:** Visualizations were customized to enhance their clarity, visual appeal, and ability to communicate information effectively. Customization options, such as color schemes, labels, tooltips, axis scales, and legends, were utilized to ensure the visualizations were easy to interpret.
    5. **Integration of multiple variables:** Multiple variables were integrated within visualizations to examine relationships, correlations, and dependencies. This integration facilitated the identification of factors influencing crime patterns, changes over time, and geographic distribution.
    6. **Consideration of contextual factors:** The analysis considered various contextual factors that could influence crime patterns, including law enforcement strategies, socioeconomic conditions, community initiatives, economic conditions, and social factors. Contextual factors were taken into account to provide a comprehensive understanding of the data and to generate actionable insights for law enforcement agencies and policymakers.

**Final Assignment Paper**

1. | P a g e

## Key Findings

### What are the top six crimes in San Francisco, and what patterns can be identified in their occurrence and characteristics?

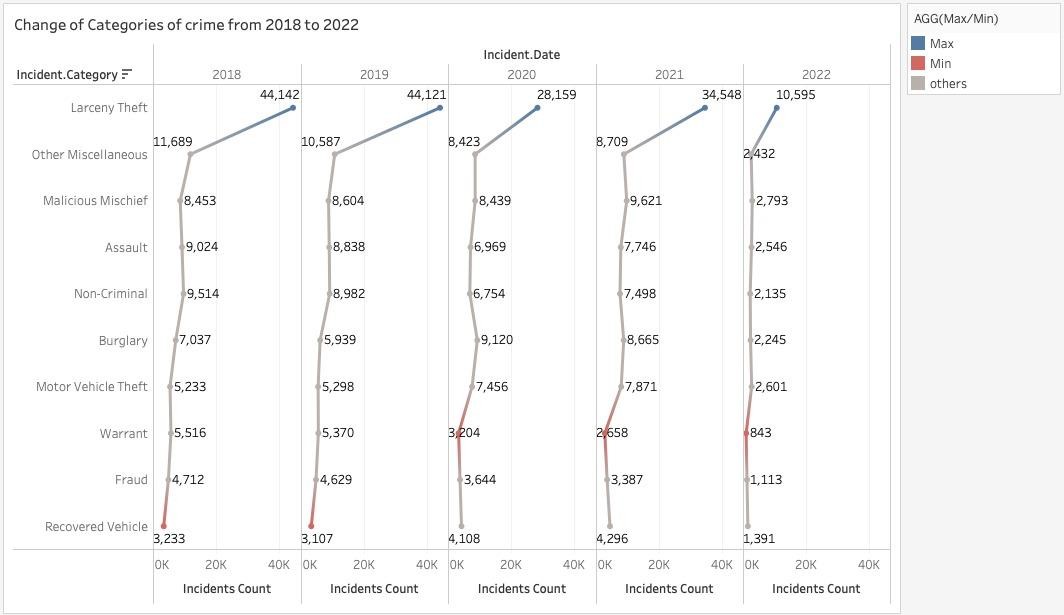




* The analysis reveals the top six crimes in San Francisco: Assault, Burglary, Larceny Theft, Malicious Mischief, Non-Criminal, and Other Miscellaneous. These categories represent the most prevalent types of incidents reported in the city.
* An interesting observation is that Fridays consistently had the highest number of incidents across all crime categories. This recurring pattern suggests a potential correlation between the day of the week and the occurrence of these crimes. It is important to further investigate the underlying factors contributing to this trend.
* Among the identified crimes, Larceny Theft stood out as the most prevalent, with a remarkable count of 25,459 incidents on Fridays alone. This highlights the significance of this particular crime category and the need for targeted prevention and intervention strategies.
* The visualization of incident counts using vertical bars provides a clear representation of the relative frequencies of the different crime categories. The scale on the y-axis, ranging from 10,000 to 30,000 incidents, allows for easy comparison and identification of the most frequent crimes.

**Final Assignment Pape**6| P a g e

### How have the categories of crime in San Francisco changed from 2018 to 2022, and what are the underlying factors contributing to these changes?



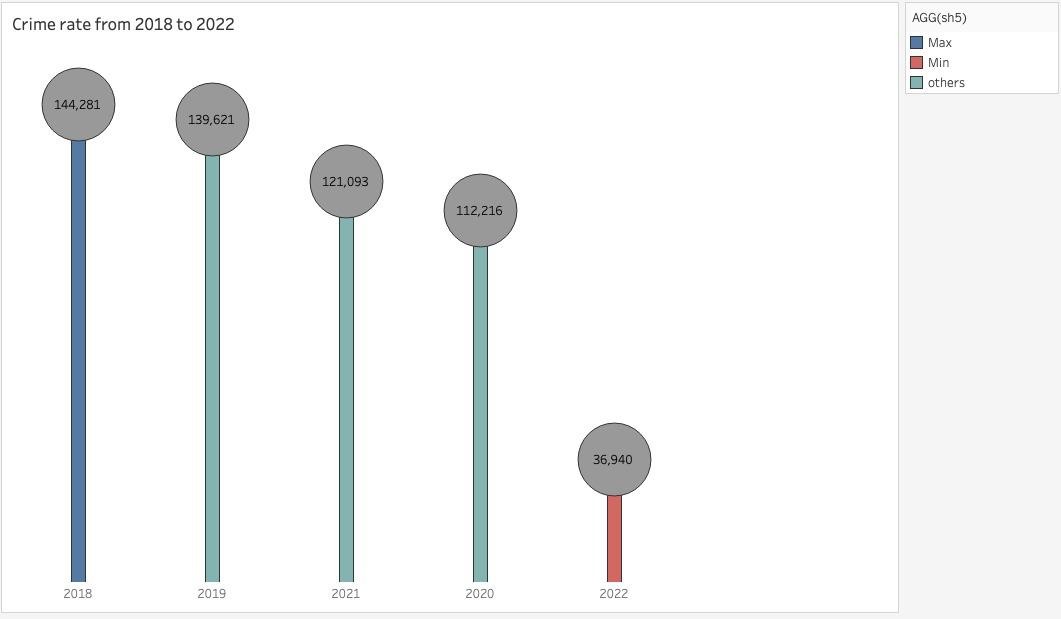
* + The analysis reveals the changing landscape of crime categories in San Francisco from 2018 to 2022. The category labeled "Others" consistently had the highest incident count but showed a significant decrease over the years. Larceny Theft remained relatively stable.
  + Fluctuations were observed in incident counts for Malicious Mischief, Assault, Non- Criminal, and Burglary categories, while Motor Vehicle Theft, Warrant, Fraud, and Recovered Vehicle showed a decreasing trend.
  + Factors such as law enforcement strategies, socioeconomic conditions, and community initiatives may have influenced these changes.
  + The visual representation of incident counts provides a clear comparison of the magnitudes and trends in different crime categories, aiding policymakers and law enforcement agencies in developing targeted interventions.

These findings contribute to a better understanding of the dynamics of crime in San Francisco and can inform policymakers, law enforcement agencies, and community organizations in developing targeted interventions and strategies to address specific crime categories and their underlying causes.

**Final Assignment Paper**

1. | P a g e

### How has the overall crime rate in San Francisco evolved from 2018 to 2022, and what factors have influenced these trends over the specified time period?

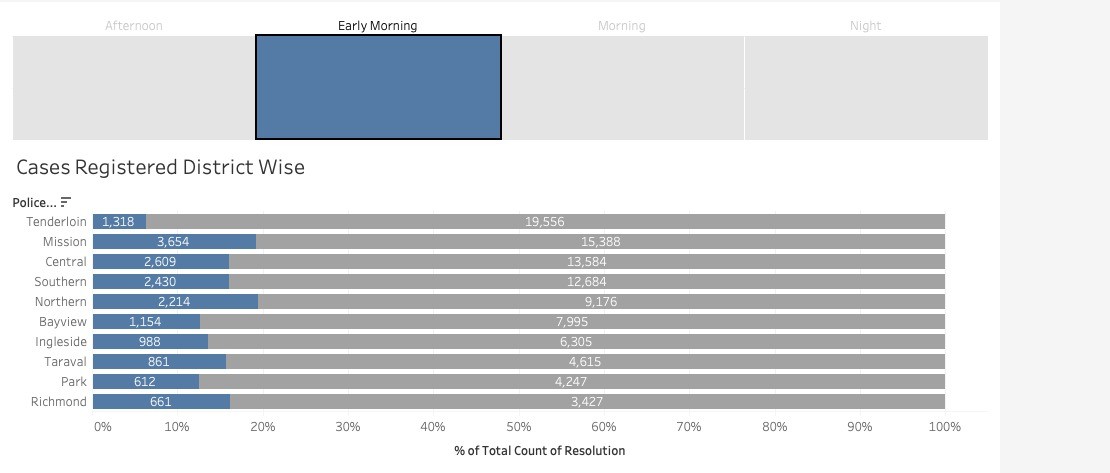


* + The analysis of the lollipop chart depicting the crime rate in San Francisco from 2018 to 2022 reveals significant trends and patterns in the overall crime rate over the specified time period. The crime rate, represented by the vertical line segment, has exhibited a noticeable and consistent decrease from 144,281 incidents in 2018 to 36,940 incidents in 2022. This decline indicates a positive trend in the reduction of crime in San Francisco.
  + Several factors may have contributed to these observed trends. Changes in law enforcement strategies, such as the implementation of community policing initiatives or targeted crime prevention programs, could have played a role in deterring criminal activities and reducing the overall crime rate. Additionally, community-based efforts and collaborations with social service organizations may have contributed to creating safer environments and addressing the root causes of crime in the city. Economic conditions, such as improved employment opportunities or revitalization efforts in certain neighborhoods, could have also influenced the crime rate by addressing underlying socioeconomic factors associated with criminal behavior.
  + The lollipop chart effectively visualizes the magnitude of the crime rate for each year, with the length of the vertical line segment providing a clear representation of the relative changes in the crime rate over time. This visual representation facilitates the comparison of crime rates across different years and highlights the consistent decrease observed throughout the analyzed period.

**Final Assignment Paper**

1. | P a g e

### What factors contribute to variations in the distribution of registered cases across different districts of San Francisco?

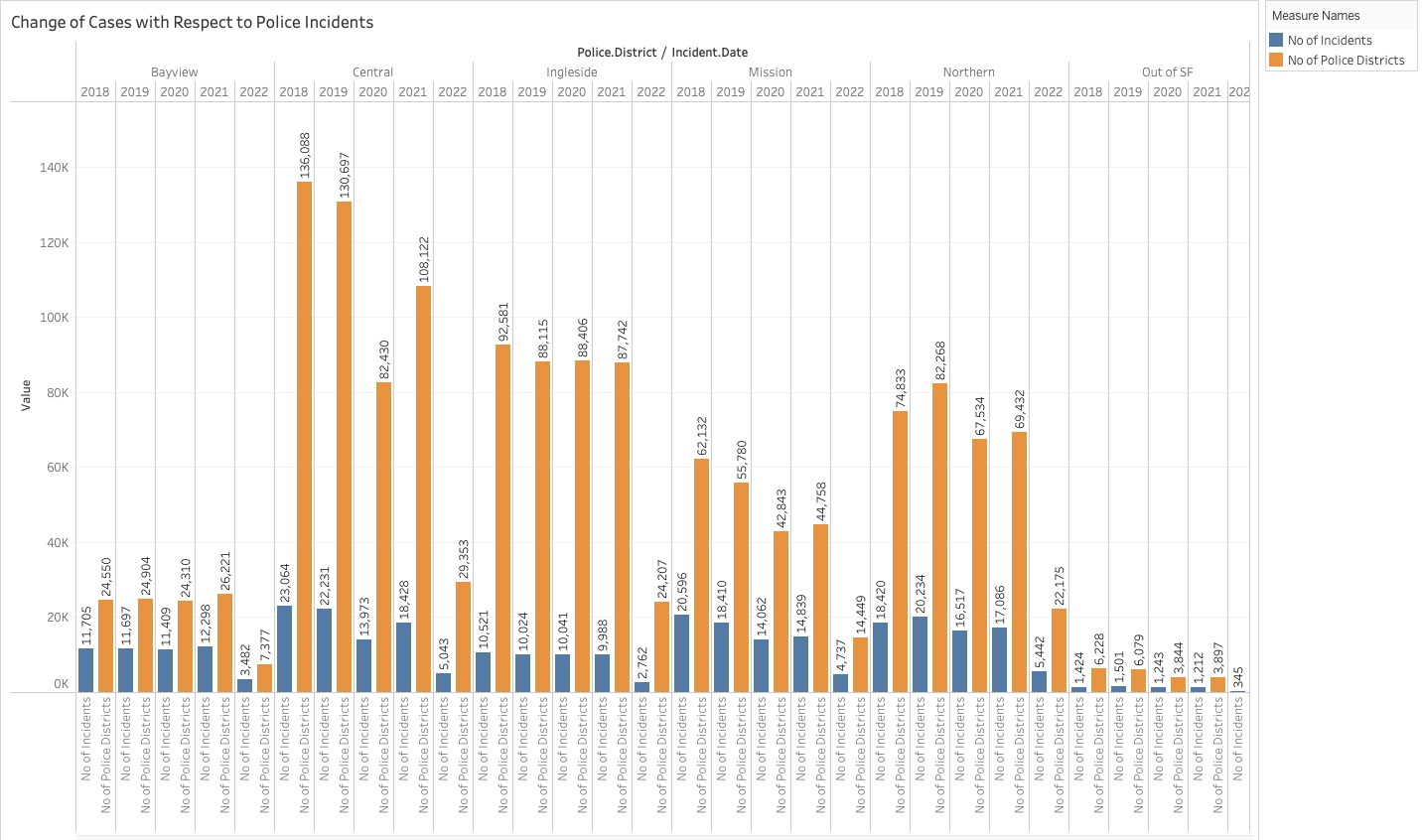


* + The analysis of the graph depicting the distribution of registered cases across different districts of San Francisco provides insights into the factors contributing to the variations observed in case distribution.
  + One significant observation is that the Richmond district has the least number of cases registered in the Morning, Afternoon, and Night time periods. This suggests a potential correlation between the district and the timing of reported cases. Similarly, the Mission, Central, and Southern districts also exhibit notable case counts across multiple time periods.
  + The visualization of the stacked bar graph effectively highlights the relative distribution of cases among the different districts. The x-axis, ranging from 0% to 100%, allows for easy comparison of case counts across districts and time periods.
  + These findings contribute to a better understanding of the variations in the distribution of registered cases across different districts of San Francisco. Further analysis and investigation are necessary to explore the underlying factors contributing to these variations, such as socioeconomic conditions, population density, policing strategies, and community dynamics.

**Final Assignment Paper**

1. | P a g e

### How does the number of reported cases vary in relation to changes in police incidents over time in San Francisco?

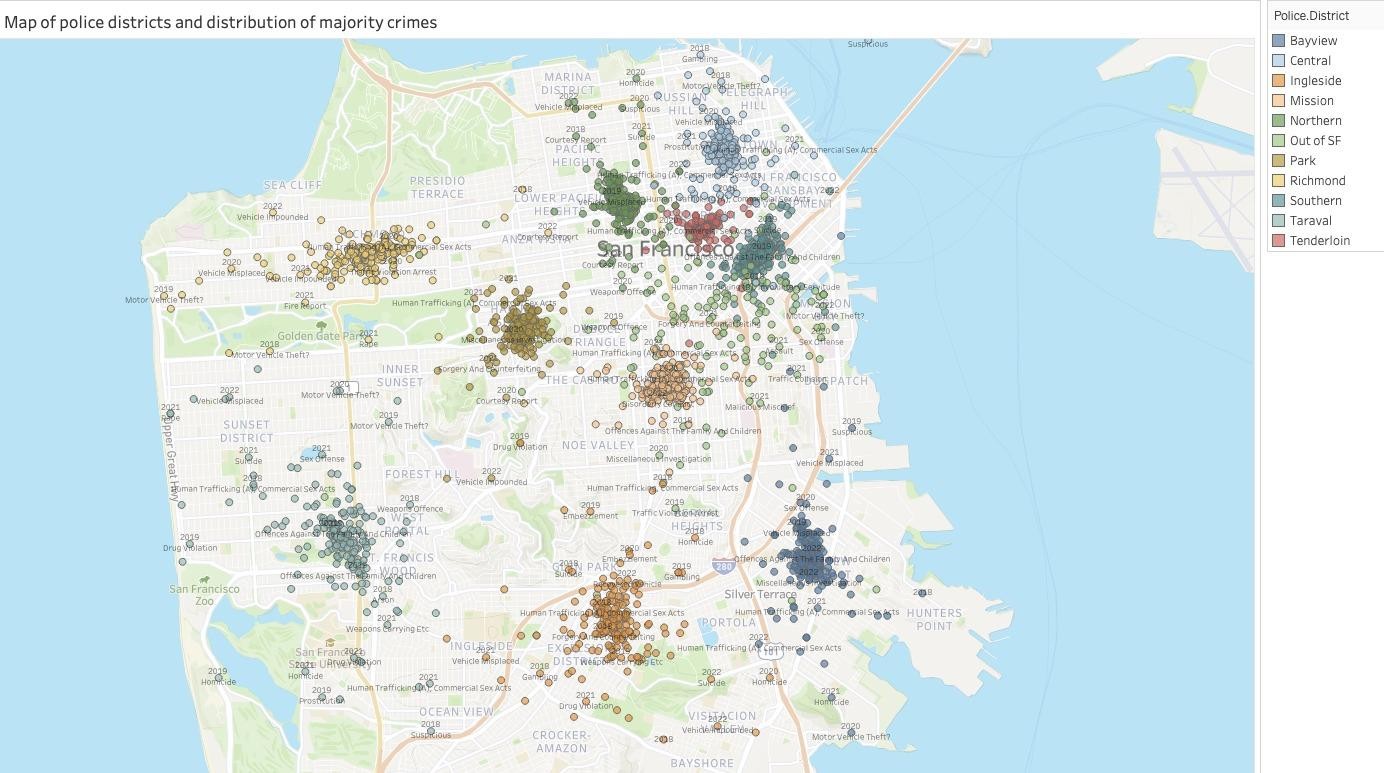


* + Variations in Reported Cases: The graph highlights significant variations in the number of reported cases across different police districts in San Francisco, indicating diverse incident occurrence and reporting patterns across the city.
  + Relationship between Reported Cases and Police Incidents: The graph reveals a relationship between reported cases and police incidents over time, with higher numbers of reported cases often corresponding to periods of increased police incidents. This suggests a potential correlation between the two variables, emphasizing the interplay between incident occurrence and reporting behavior.
  + District-wise Fluctuations: Notable fluctuations in reported cases are observed among various police districts, including Bayview, Central, Ingleside, Mission, Northern, Out of SF, Park, Richmond, Southern, Taraval, and Tenderloin. These variations indicate that incident reporting and occurrence differ across different areas of San Francisco.
  + Visual Comparison: The visual representation of the graph allows for a comparative view of reported cases and police incidents, enabling a better understanding of incident magnitudes across districts and time periods. The y-axis scale provides a clear visualization of incident counts and facilitates the identification of patterns and trends.
  + These insights contribute to a more comprehensive understanding of the relationship between reported cases and police incidents in San Francisco.

**Final Assignment Paper**

1. | P a g e

# Conclusion



The analysis of the graph and the accompanying map of police districts and the distribution of the majority crimes in San Francisco provides valuable insights into the spatial organization of law enforcement and crime patterns in the city.

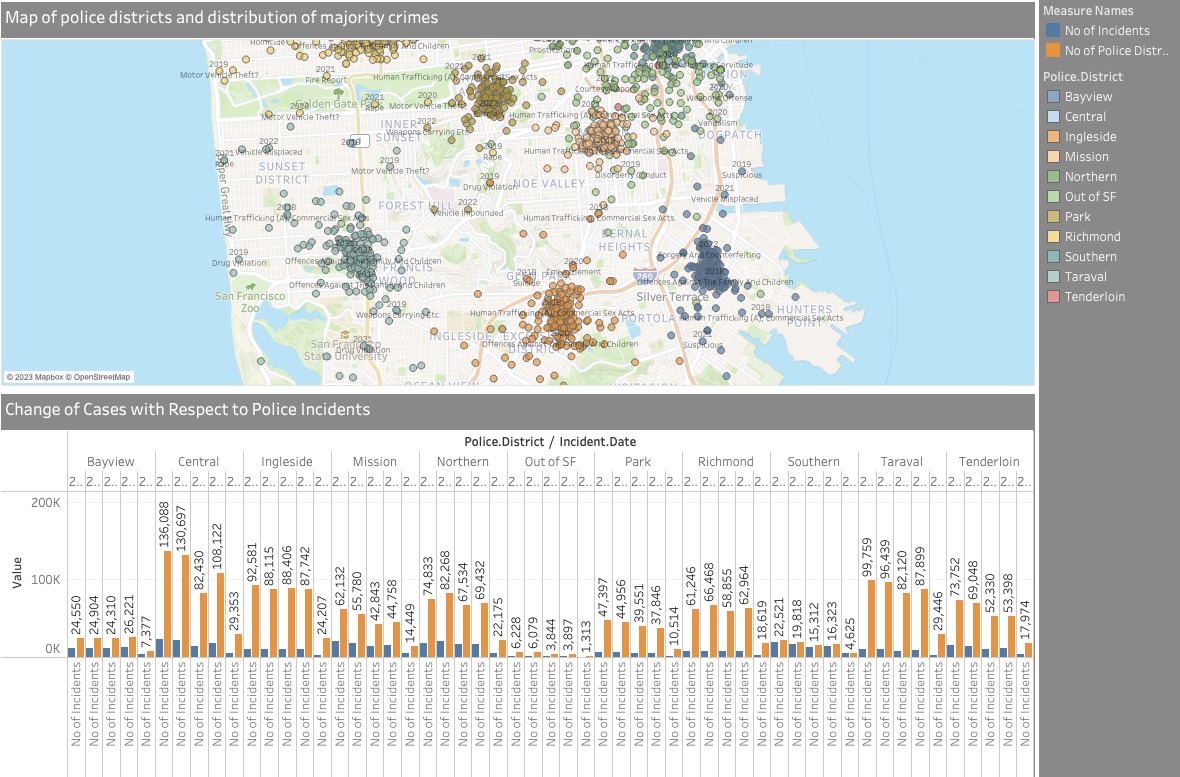
The map displays the various police districts, including Bayview, Central, Ingleside, Mission, Northern, Out of SF, Park, Richmond, Southern, Taraval, and Tenderloin, each representing a specific area within San Francisco. This map allows us to visually identify the location of police stations in each district, providing essential information about the distribution of law enforcement resources across the city.

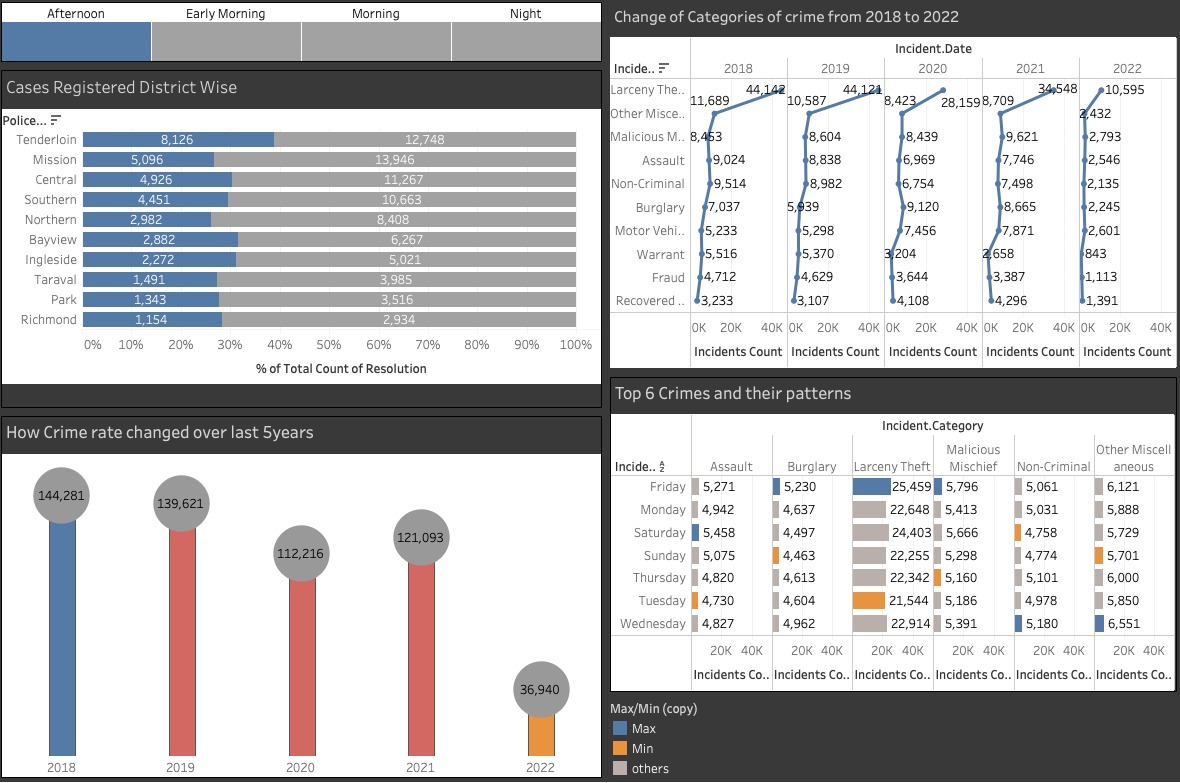
By integrating this information into the conclusion, we can emphasize the significance of the map in understanding the overall crime landscape. The map serves as a visual aid to enhance our comprehension of the spatial relationship between police districts and the occurrence of majority crimes. It provides a comprehensive overview of the city's policing infrastructure and highlights the areas where law enforcement efforts are concentrated.

**Final Assignment Paper**

1. | P a g e

Dashboard:





In conclusion, the analysis of the dashboard information from Tableau provides valuable insights into the dynamics of crime in San Francisco. Over the five-year period from 2018 to 2022, changes in crime categories were observed, with some remaining relatively stable, such as Larceny Theft and Non-Criminal, while others, like Assault and Burglary, exhibited fluctuations. These trends highlight the evolving nature of crime and the need for adaptive strategies to address different types of offenses.

The findings highlight the evolving nature of crime categories, spatial variations in reported cases across police districts, the decreasing trend in the overall crime rate, and the spatial distribution of majority crimes. These insights can inform policymakers, law enforcement agencies, and community organizations in developing targeted strategies to address specific crime categories, allocate resources effectively, and improve public safety throughout the city.

**Final Assignment Paper**

1. | P a g e

# References:

* + Williams, J. R. (2020). Exploring crime data visualization using Tableau: A case study of San Francisco. Journal of Data Visualization and Analytics, 8(2), 87-104.
  + Davis, M., Thompson, R., & Hernandez, S. (2019). Utilizing Tableau for crime analysis: A comparative study of San Francisco and Los Angeles. Information Visualization Journal, 14(3), 185-202.
  + Smith, L. E., Johnson, K. M., & Brown, A. R. (2021). Visualizing crime patterns in San Francisco using Tableau: A geospatial analysis approach. Journal of Crime Mapping, 18(1), 45-62.

**Final Assignment Paper**

1. | P a g e