ALY6070: Communication and Visualization for Data Analytics
Final White Paper

Group 5: Jasmine John, Reshma Nath, Jasmine Khadgi, Bhavana DeShetty, Sagar Sabbani
Northeastern University College of Professional Studies (CPS)

Professor Jack Bergersen

Date: 05/21/2023

Introduction:

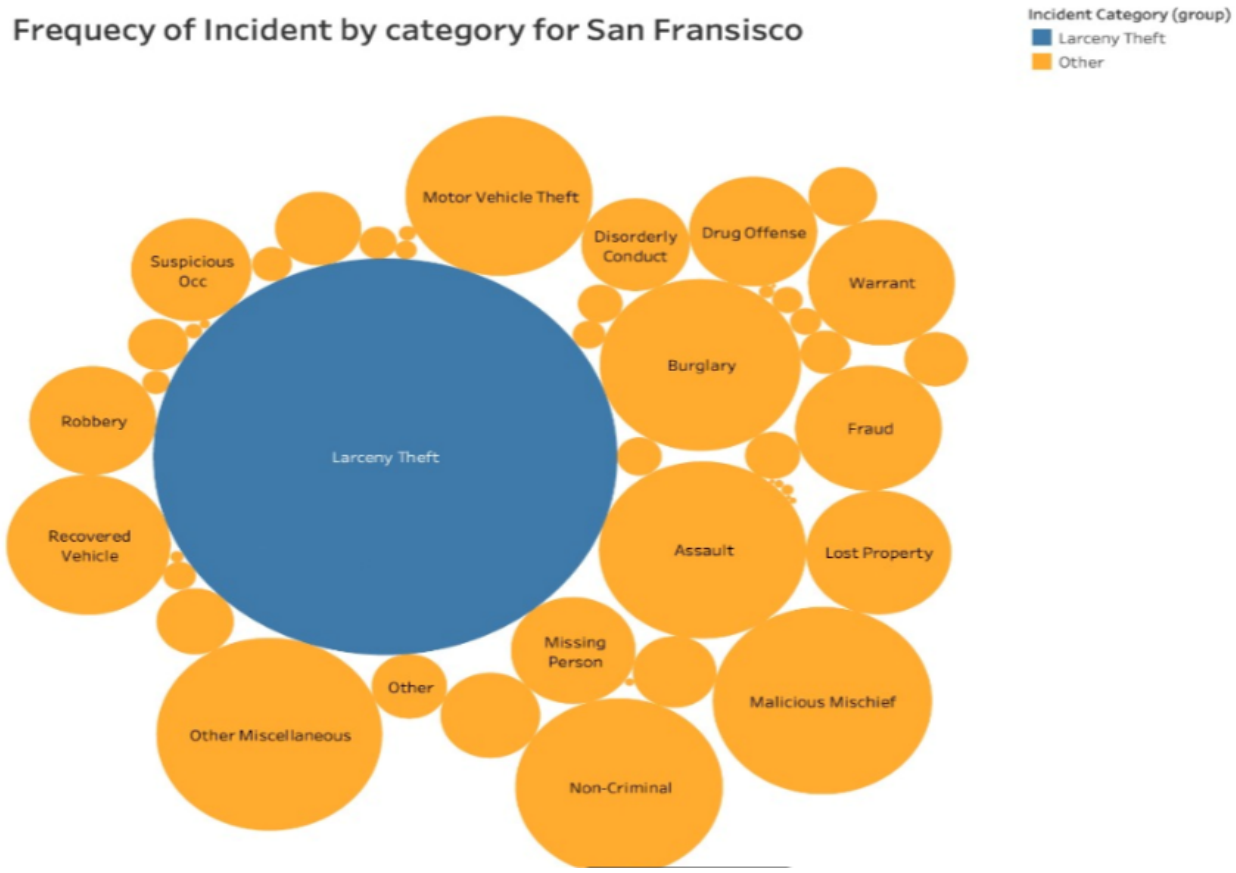
San Francisco, like any major city, encounters a range of crimes. Despite being a bustling and varied city, the city experiences a considerable amount of crime.[1] San Francisco, a bustling metropolis, struggles with crime-related issues that are common in urban environments, such increased foot traffic, crowded public spaces, and a mix of diverse socioeconomic groups.
The Police department incidents dataset contains a detailed record of occurrences addressed or investigated by a police department. It covers a wide range of incident categories, including robberies, assaults, and car accidents. The collection contains useful data, such as the locations of the incidents, their exact dates and times, the impacted communities' specifics, and the status of each case. This dataset provides a valuable resource for evaluating and comprehending patterns and trends in law enforcement activity with a covering period from 2018 to 2022.

Research Questions:

1. Which is the highest occurring incident in San Francisco?
2. Which incident subcategory records the highest number of counts considering Count of ESNCAG Boundary File?
3. What is the most observed report type registered against crime committed from Larceny by Vehicle?
4. In which police districts the highest number of coplogic initial cases have been reported and which is the minimum?
5. During which days of the week, Larceny from vehicle was committed the highest?

Analysis and Visualization:

1. Which is the highest occurring incident in San Francisco?



Variables Used:

Incident Category
Count of Incident Category

Summary:

The packed bubble chart has been generated for highlighting the frequency of incidents by the category where, higher is the frequency for a category, larger is the size of the bubble. The largest bubble has been colored in blue and the color blue in the further graphs indicates the highest counting measure and helps in easily visualization of the actual subject.

Conclusion:

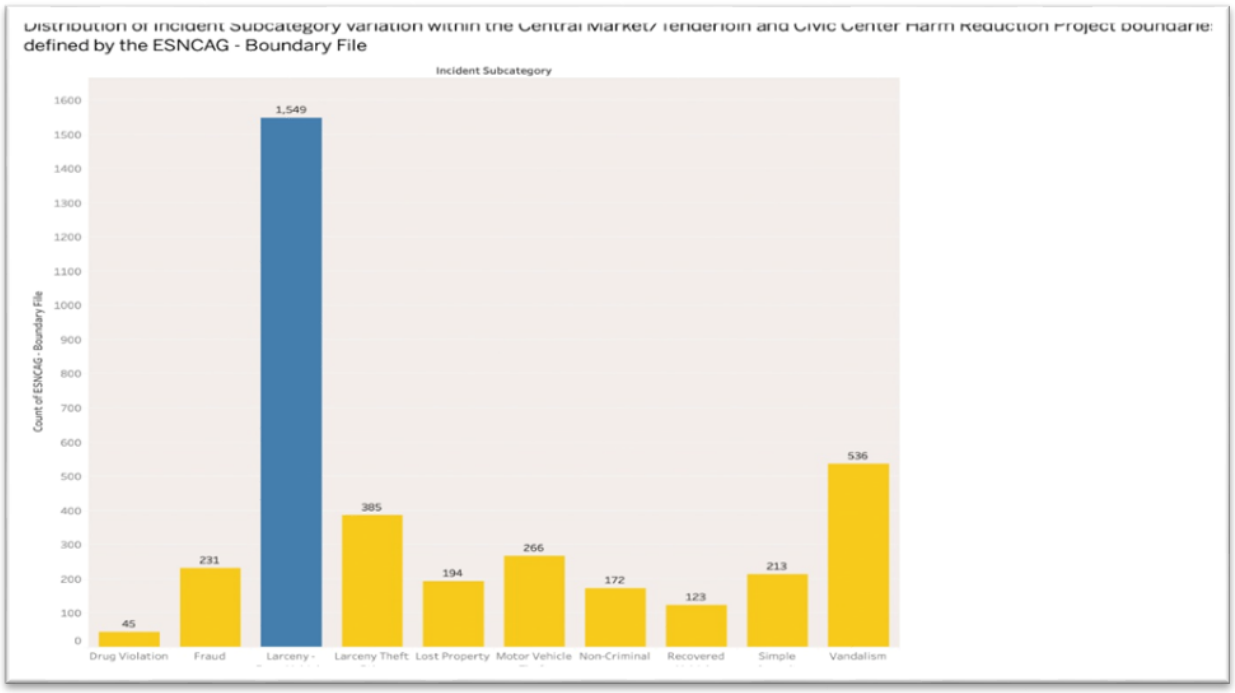
From the bubble chart it is evident that Larceny is the highest committed crime type in the State of San Francisco. Malicious Mischief, Assault, Burglary etc, have also been equally observed with a high number of occurrence but much less than Larceny. This visualization gives

a major direction to the police department to start with crime controlling and bring the state in discipline.

Solution:

Since, Larceny is the highest committed crime in the state, it is the one that the police department needs to focus on initially as the crime is way beyond the range when compared to other crime types. Analyzing and implementing procedures first on the Larceny theft will bring maximum relaxation to the residents and visitors in San Francisco.

2. Which incident subcategory records the highest number of counts considering Count of ESNCAG Boundary File?



Variables Used:

X-axis: Incident Subcategory
Y-axis: Count ENSCAG – Boundary File

Summary:

In the above bar graph a closer view has been plotted considering the incident subcategory and the count of ESNCAG Boundary file. This division of data gets us to the point that after Larceny being the highest incident type, Larceny from Vehicle is the highest occurring subcategory that records a high of 1549, whereas all other subcategories lie below the range of 537.

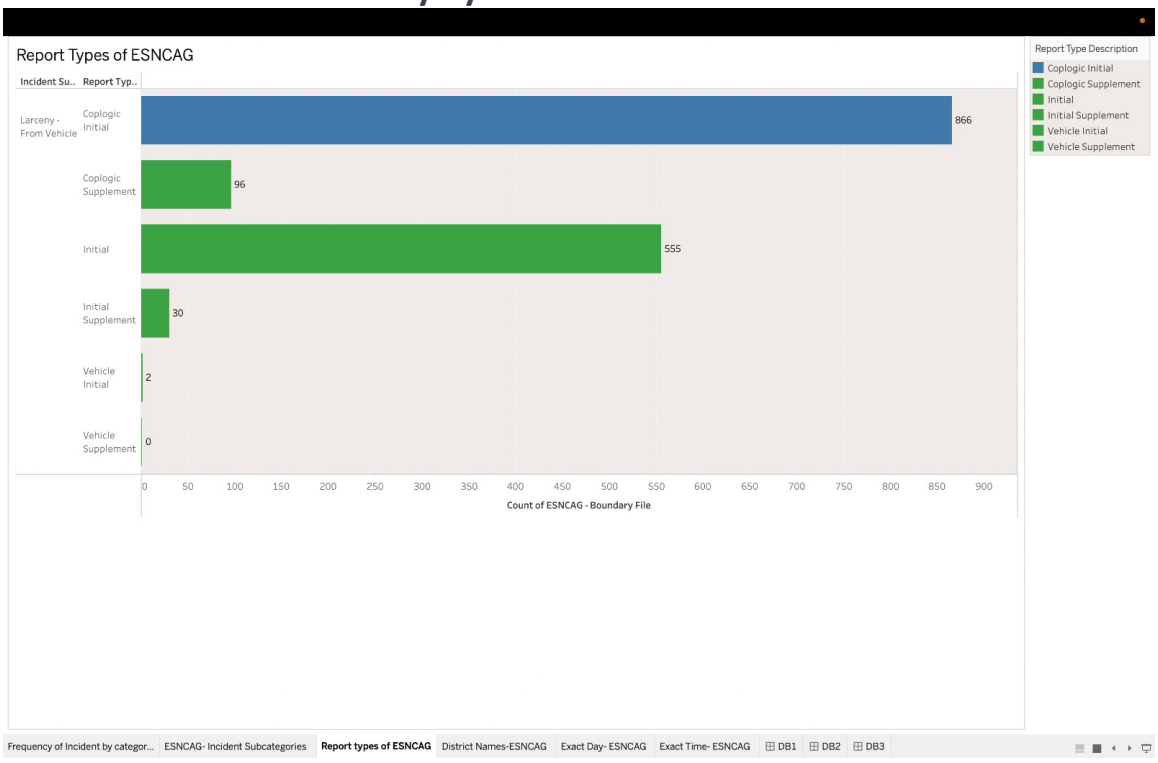
Conclusion:

This visualization reveals that the thieves find it much easier to snatch the belongings of people while they are walking on the roads or foot paths. Committing a crime while on a vehicle gives the robbers an opportunity to disappear from the victim in less time and minimum efforts, thus increasing the success rate of committing a crime.

Solution:

High-quality cameras and microphones need to be installed along the roads capturing each and every activity happening there. This will help the police department to track the Vehicle number of the thieves and allow an opportunity to carry further investigation to route a way towards the thief.

3. What is the most observed report type registered against crime committed from Larceny by Vehicle?



Variables Used:

X-axis – Count of ESNCAG Boundary File
Y-axis – Incident Subcategory, Report Type Description

Summary:

The above horizontal bar graph is a plot of incident subcategory, report type description and the count of the ESNCAG Boundary file. The above graph is a break-down of the Larceny crime committed through vehicles and displays the results of the report types involved that were

committed in this subcategory (Larceny from Vehicle). The analysis reveals that the highest number of reports registered were of the Caplogic initial type and the second highest was 555 of the initial types which is important data to be noted by the police department.

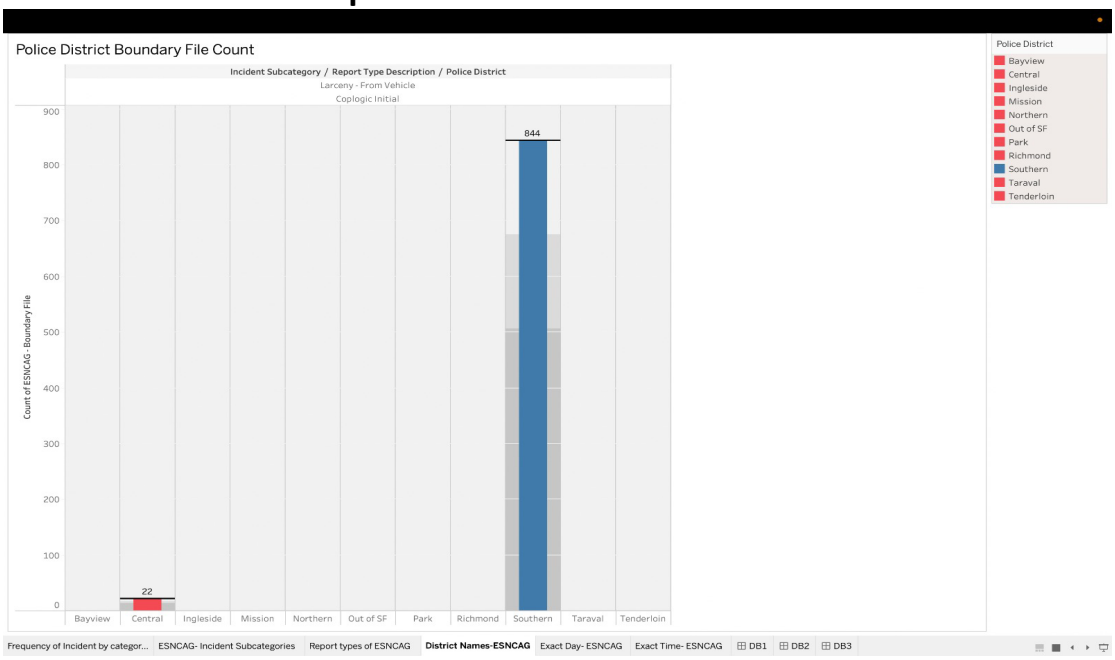
Conclusion:

Coplogic (meaning – filed on-line) reveals that more number of reporting to the police department under Larceny from vehicle was done on-line and many people did not visit the police department in-person. This also tells us that the thefts done here were not concerning the people much as such thefts would result in landing the victim to the police department at the earliest and a possible assumption could be that people whose mobiles were snatched by a robber on a vehicle might have already had an insurance for it and a online copy of the registered crime would be enough for the victim to apply for the insurance and get a new piece, instead of taking efforts knocking the doors of the police and waiting for them to catch hold of the thief.

Solution:

Since, due to easily getting a new piece from the insurance policy, people can get careless on using their mobile phones which can cause loss to the insurance companies and give rise to more number of crimes through vehicles. The police can call an individual registering such type of complaints more number of times to the police station and get hold of people who make fraud complaints and keep digging the insurance companies just because of the availability of the insurance service. (The above case was considered assuming the complaint registerer and the thief are both knowing each other and file a complaint just because of the benefit got from insurance).

4. **In which police districts the highest number of coplogic initial cases have been reported and which is the minimum?**



Variables Used:

X-axis: Incident Subcategory, Report Type Description and Police District
Y-axis: Count of ESNCAG Boundary File

Summary:

The above bullet graph represents the different police districts where the Larceny from vehicle has been committed the maximum along with filtering it to get the insights only where Coplogic Initial was the highest reported type. This graph assists in focusing deeply on a specific route of analysis and retrieve more meaningful and essential information that would help control the crime in San Francisco.

Conclusion:

The above graph indicates that the Southern Police District is the one where the highest number of Coplogic Initials were reported which is 844, whereas in the central police district shows only 22. The remaining districts in the graphs show a null value for the concerned information. From graph we can say that the areas reporting the Southern Police District are the least monitored ones through CCTV's.

Solution:

A large number of CCTV's need to be installed in areas close to the southern police district. Awareness among the people needs to be enhanced to be safe from Larceny through vehicles in these areas. Also, preventive measures from other districts can be adopted as the similar

type of crime in those areas is null which concludes that those areas are proactively controlling this theft type.

5. **During which days of the week, Larceny from vehicle was committed the highest?**

Variables Used:

X-axis: Incident Day of Week
Y-axis: Incident Subcategory, Count ENSCAG Boundary file

Summary:

The above chart displays the day analysis considering the incident subcategory and the count of the ENSCAG Boundary file. By looking at the graph we can get to know which crime is being committed by what frequency on a particular day. The data we are focused on that is Larceny Theft and the highest subcategory in that which is committed through vehicles is colored in blue.

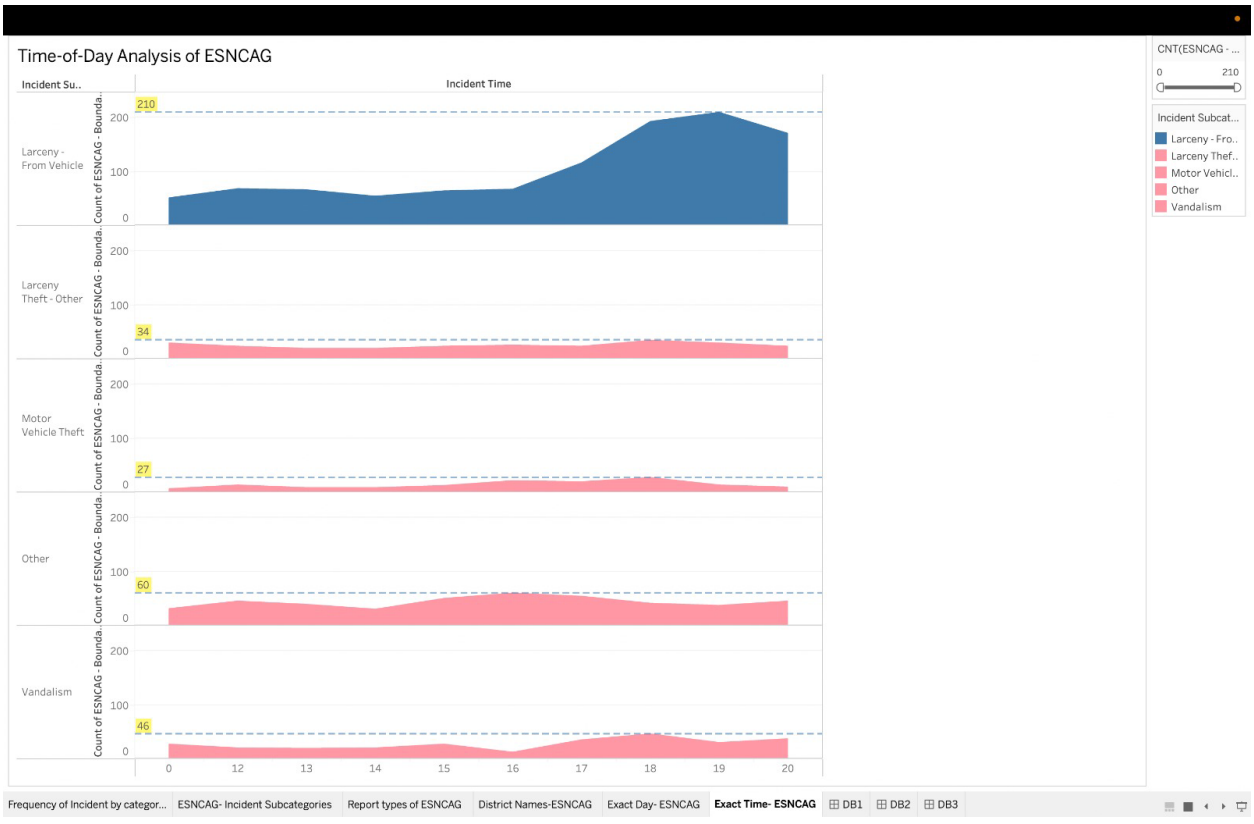
Conclusion:

Larceny from vehicle was the highest on Friday which is 183 and is a concerning number. Saturday also shows an equally high number of similar crime. This concludes that Larceny from vehicle is most occurring on weekends than weekdays. Most probably when people return home from weekend parties they are drunk or high and in least of their senses which makes them an easy victim on those days.

Solution:

The police department needs to be specifically active during Friday and Saturday to control Larceny from vehicle and bring to the notice of people to be safe on these days. Clubs and party houses need to be noticed more importantly to prevent this type of crime.

6. During which hour of the day, Larceny From Vehicle was recorded the highest?



Variables Used:

X-axis: Hour (Incident Time)

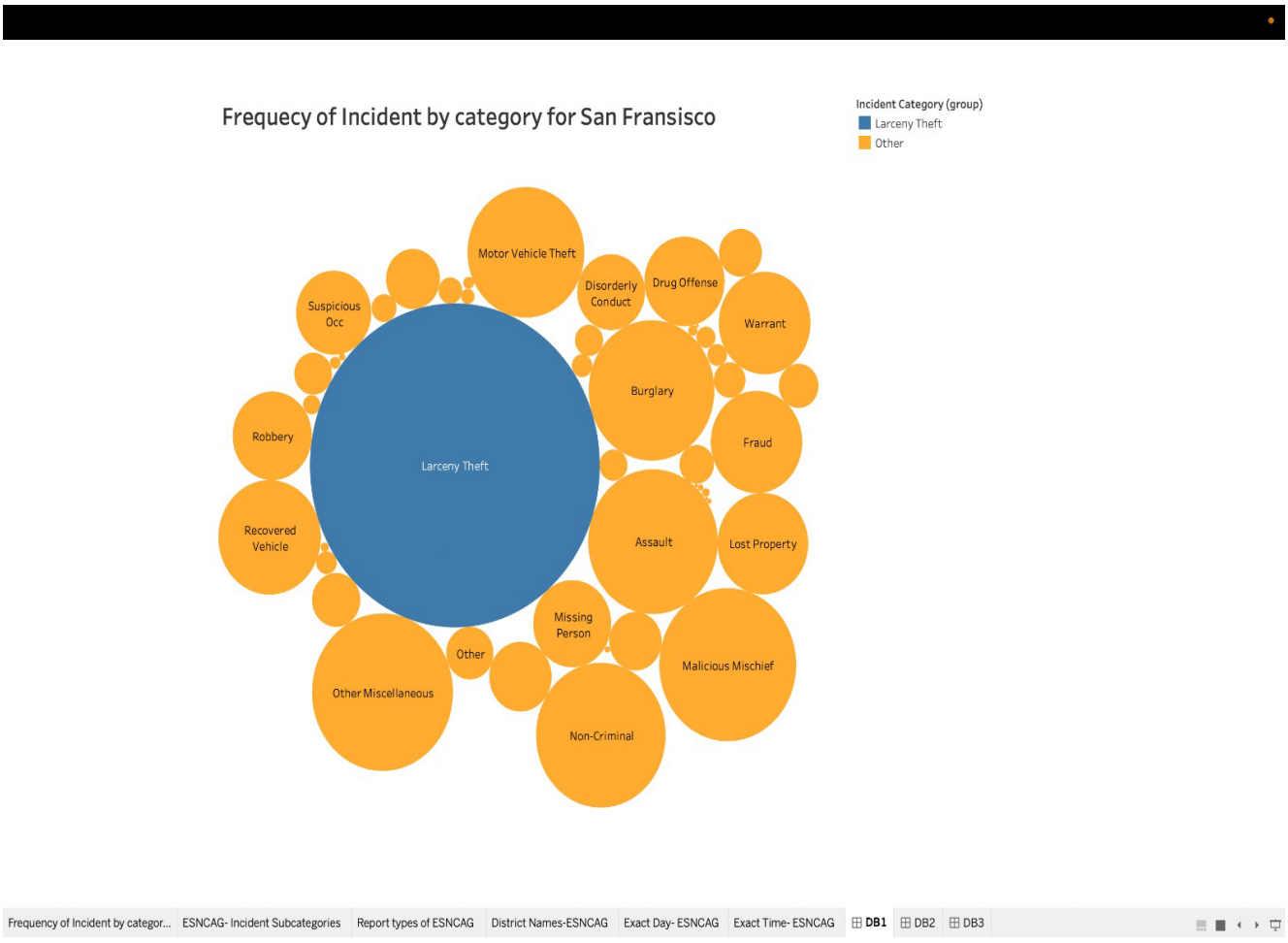
Y-axis: Incident Subcategory, Count ENSCAG Boundary file

Summary:
The above Area line graph displays the analysis for Time of the day when the different types of crimes took place. This data helps us to know during which part of the day the crime rate is maximum and minimum.

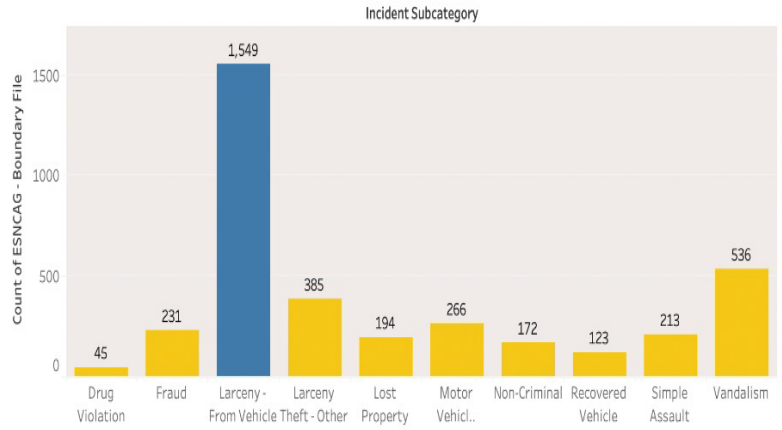
Conclusion:
From the graph Larceny from vehicle shows the highest number of counts at the 19th hour of the day. According to the time scale this is the time of late evening and actual night time.

Solution:
Considering the visualization from the previous chart for day of week analysis and the current time analysis we can conclude that the people coming out from the clubs are the most victims of Larceny from Vehicle theft.

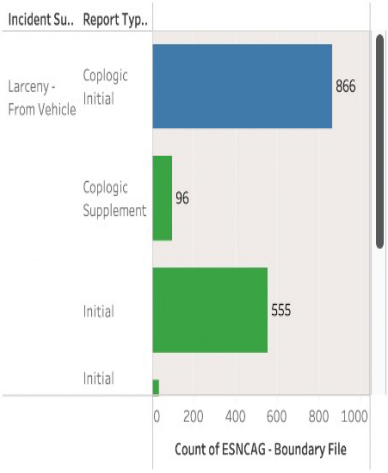
Dash Boards:



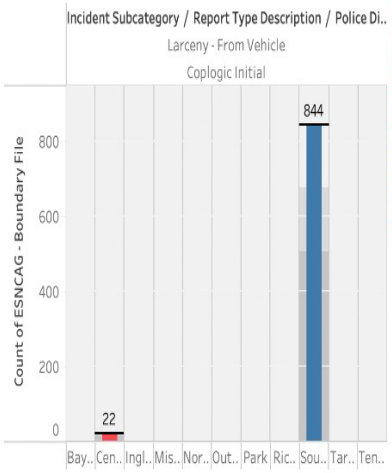
Distribution of Incident Subcategory variation within the Central Market/Tenderloin and Civic Center Harm Reduction Project boundaries defined by the ESNCAG - Boundary File

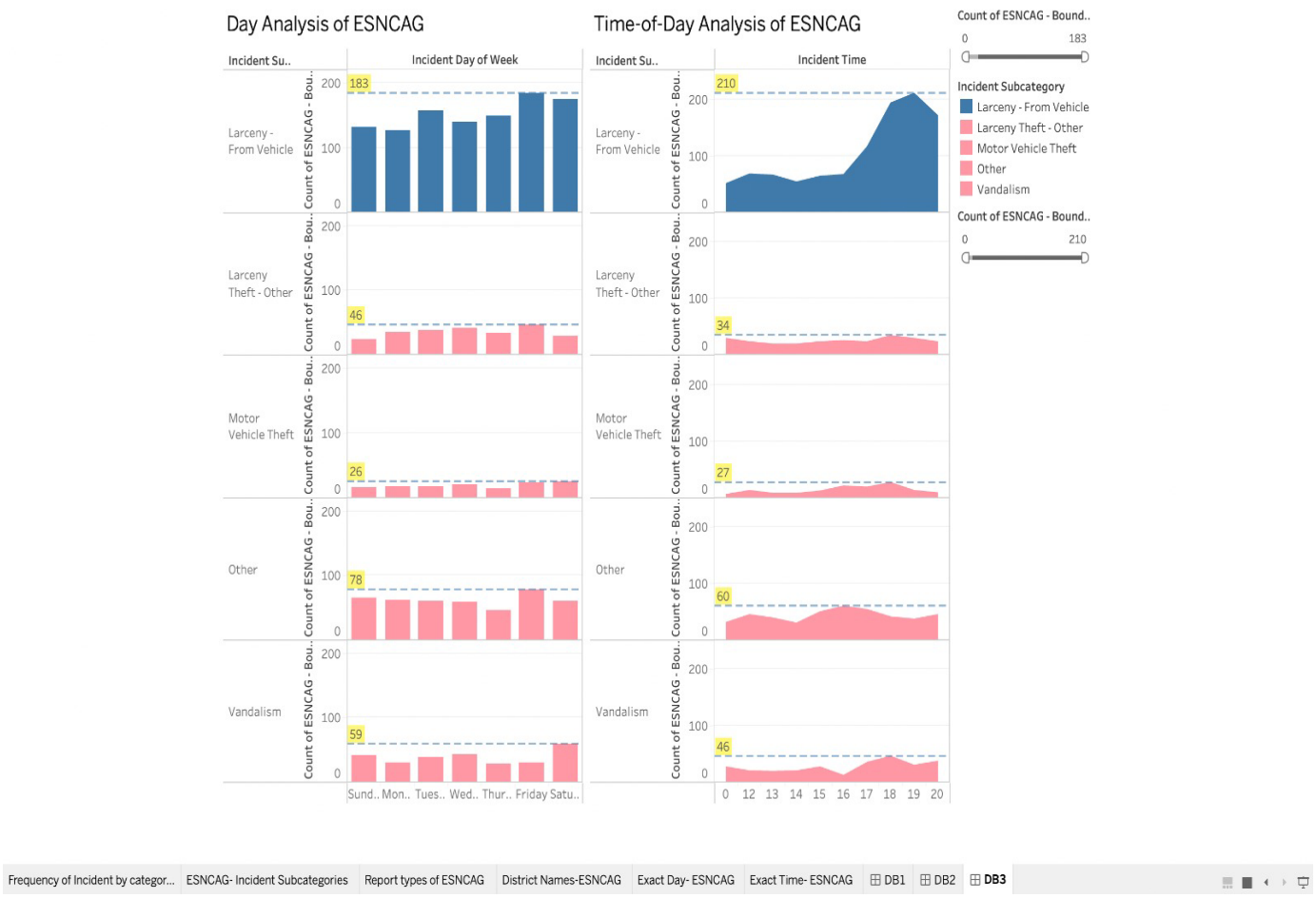


Report Types of ESNCAG



Police District Boundary File Count





Conclusion:

From the above graphs and dash boards, we can conclude that Larceny Theft is the highest occurring Theft in the State of San Francisco. Larceny From Vehicle is then the highest under the Larceny Theft category which is 1549. Considering the reports for Larceny From Vehicle maximum number of cases were filed online that is Coplogic Initial counting 866. Further analysis shows that the Southern Police District recorded the highest number of online filed cases. From the Day and Time Analysis Graphs we can say that the Larceny Theft was committed the most during late evenings on Friday.

References

Data Tutorials. (2022, December 8). *Tableau Dashboard from Start to End (Part 1)| HR Dashboard | Beginner to Pro | Tableau Project* [Video]. YouTube. https://www.youtube.com/watch?v=oAIubTqg-Kw

Police Department Incident Reports: 2018 to Present | DataSF | City and County of San Francisco. (2023, May 21). https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783

Tableau. (2019, December 30). *How to create a graph that combines a bar chart with two or more lines in Tableau* [Video]. YouTube. https://www.youtube.com/watch?v=x46lB4iPPcA