San Francisco Incident Report Analysis

Jeffrey Teng

Identify Business Task:

1. What are the incidents that occurred the most within the four years period?
2. Where are the Top 8 neighborhoods that most of the incidents occurred?
3. How does the months affect the incidents?
4. Are the incidents more of emergency or not?
5. When and where will be better to increase the number of police officer to improve the safety?
6. Should city increase more funds and officer?

Prepare Phase:

The data used for this project was collected and published by the San Francisco Open Data – DataSF, the official San Francisco city and county data portal. The data contains the incident report from January 2018 to July 2021.

The data contained the following field:

Incident_Datetime: the date and time when incident occurred
Incident_Date: date when incident occurred
Incident_Time: time when incident occurred
Incident_Year: year when incident occurred
Incident_Day_of_Week: day of week of incident occurred
Report_Datetime: date and time when incident was reported
Report_Type_Code: system code of report type
Report_Type_Description: description of the report type code
Filed_Online: is the incident report online or not
Incident_Code: system code of incident
Incident_Category: category of the incident
Incident_Subcategory: subcategory of the incident
Incident_Description: detailed description of the incident
Resolution: the resolution of the incident at the time of the report
Intersection: the closet intersection of where the incident occurred
CNN: unique identifier of the intersection for reference
Police_District: police district where the incident occurred
Analysis_Neighborhood: neighborhood where the incident occurred
Supervisor_District: board of supervisor
Latitude: latitude where the incident occurred
Longitude: longitude where the incident occurred
Current_Supervisor_Districts: current supervisor district
Current_Police_Districts: current police district

Process Phase:

Tool used for the analysis:

Excel: Used this software for initial data cleaning and processing such as changing the date format to YYYY:MM:DD HH:MM and filling the empty cells with 0 in order to import csv file into MySQL workbench

MySQL: Used this to import, clean, process, extract, and export data. Also used it to analyze the top 10 most common incidents occurred, top 8 neighborhoods where the incident occurred, which month occurred more incidents than others, and emergency and non-emergency incidents.

Tableau: Used this to create multiple visualizations from the csv file exported from MySQL Workbench. Also created a dashboard of the project using this software.

Analyze Phase:

Incidents base on each years and months from January 2018 to July 2021

```
-- Count of incident base on year and month
SELECT
    incident_year AS Year,
    MONTHNAME(incident_date) AS Month,
    COUNT(incident_Description) Number_of_Times
FROM
    main_data
GROUP BY incident_year , Month
ORDER BY COUNT(incident_Description) DESC;
```

Year	Month	Number of Times
2018	August	13767
2018	July	13694
2019	August	13643
2019	October	13563
2018	January	13245

For the incidents based on year and month from January 2018 to July 2021, 2018 and 2019 are the years with the most incidents occurred. from the result, we can see that August of 2018 and August of 2019 are both in top 5 of number of incidents occurred.

Incidents base on each month from January 2018 to July 2021

```
-- Count of incident base on month
SELECT
    MONTHNAME(incident_date) AS Month,
    COUNT(incident_subcategory) Number_of_Times
FROM
    main_data
GROUP BY Month
ORDER BY COUNT(incident_subcategory) DESC;
```

As for the incidents occurred based on month, top five months are January, July, May, June, and February from 2018 to 2021. According to this data, if the SF police department want to consider which month that they should add more police force into the street, these months will be a better choice.

Incidents based on day of week from January 2018 to July 2021

```
-- Count of incident base on day of week
SELECT
    incident_day_of_week, COUNT(incident_code)
FROM
    main_data
GROUP BY incident_day_of_week
ORDER BY COUNT(incident_code) DESC;
```

From the data, the ranking are as follows, Friday, Wednesday, Monday, Saturday, Thursday, Tuesday, and Sunday. The number of times where the incidents occurred don't really have that much of difference between ran one and rank 7, they all have total of more than 60,000 incidents occurred each day of the week for the past 4 years.

Incident base on the latitude and longitude from January 2018 to July 2021

```sql
-- Incident base on lat and lng group
SELECT
    incident_category,
    COUNT(incident_Category),
    Analysis_Neighborhood,
    latitude,
    longitude
FROM
    main_data
WHERE
    latitude != '0' AND longitude != '0'
GROUP BY latitude , longitude
ORDER BY COUNT(Incident_Category) DESC
LIMIT 8;
```

Neighborhood	Number of Times
Financial District/South Beach	2855
Tenderloin	2562
South of Market	2306
Mission	2056
Tenderloin	1852

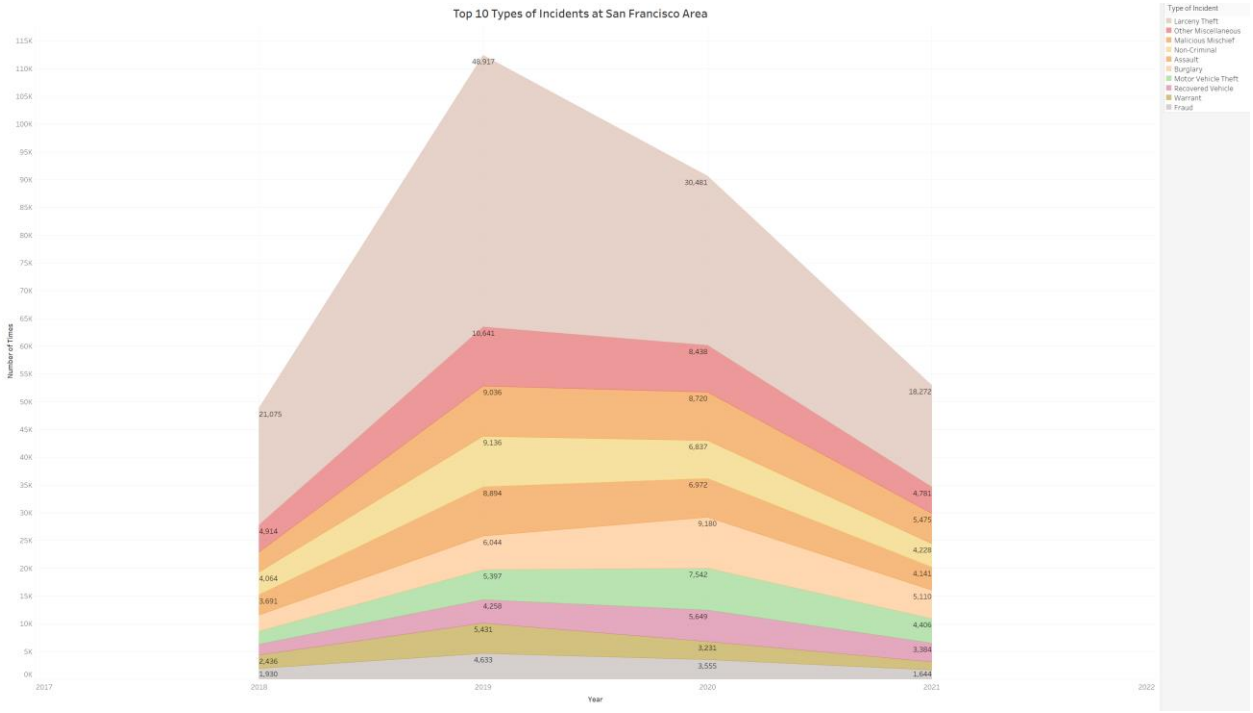
From the data above, the top 5 neighborhood where the most incidents occurred are financial district, Tenderloin, South of Market, Mission, and Tenderloin. Tenderloin showed up twice in the tables because the data was based on the latitude and the longitude. These are the neighborhood that incidents occurred the most.

Incidents that are filed online or not.

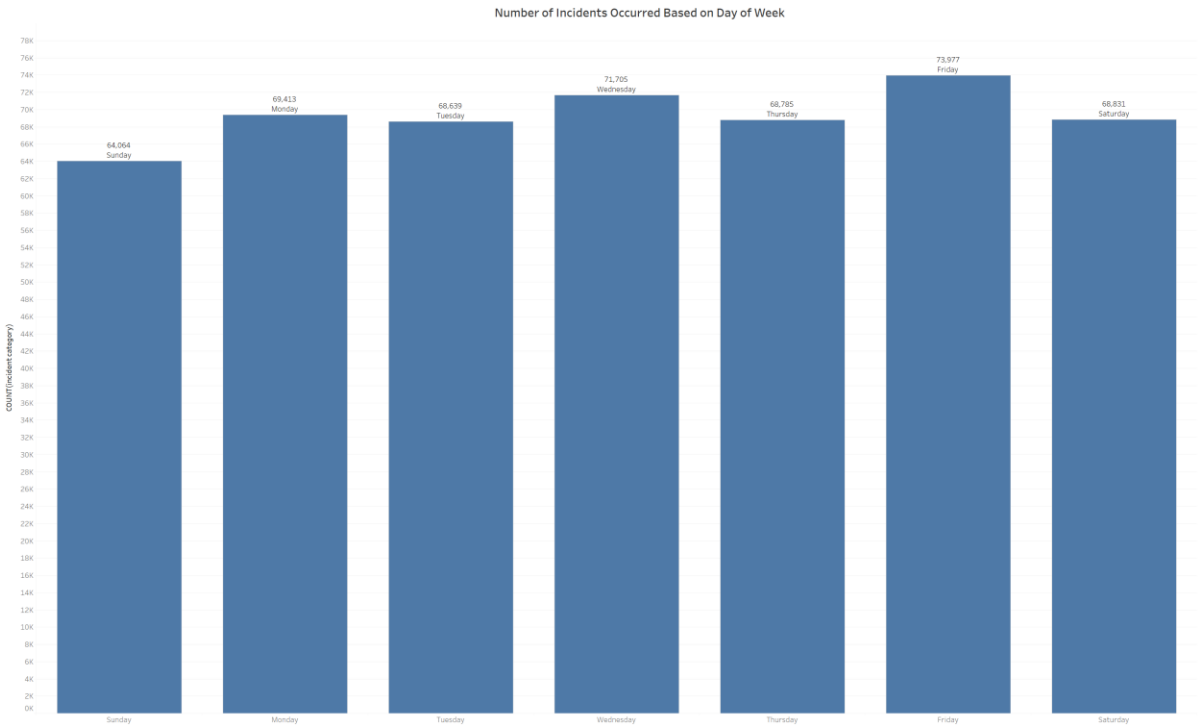
```sql
-- File online or not
SELECT
    filed_online,
    ROUND((COUNT(filed_online) / (SELECT
                    COUNT(filed_online)
                FROM
                    main_data) * 100),
            2) AS Percentage_of_File
FROM
    main_data
GROUP BY filed_online;
```

The data shown if the incidents are emergency or non-emergency, if it is non-emergency, the people can file report online but if it is an emergency, people need to call 911 as soon as possible. In this case, 19.98 percent of the report are file online which mean is non-emergency and 80.02 percent of incidents are emergency.

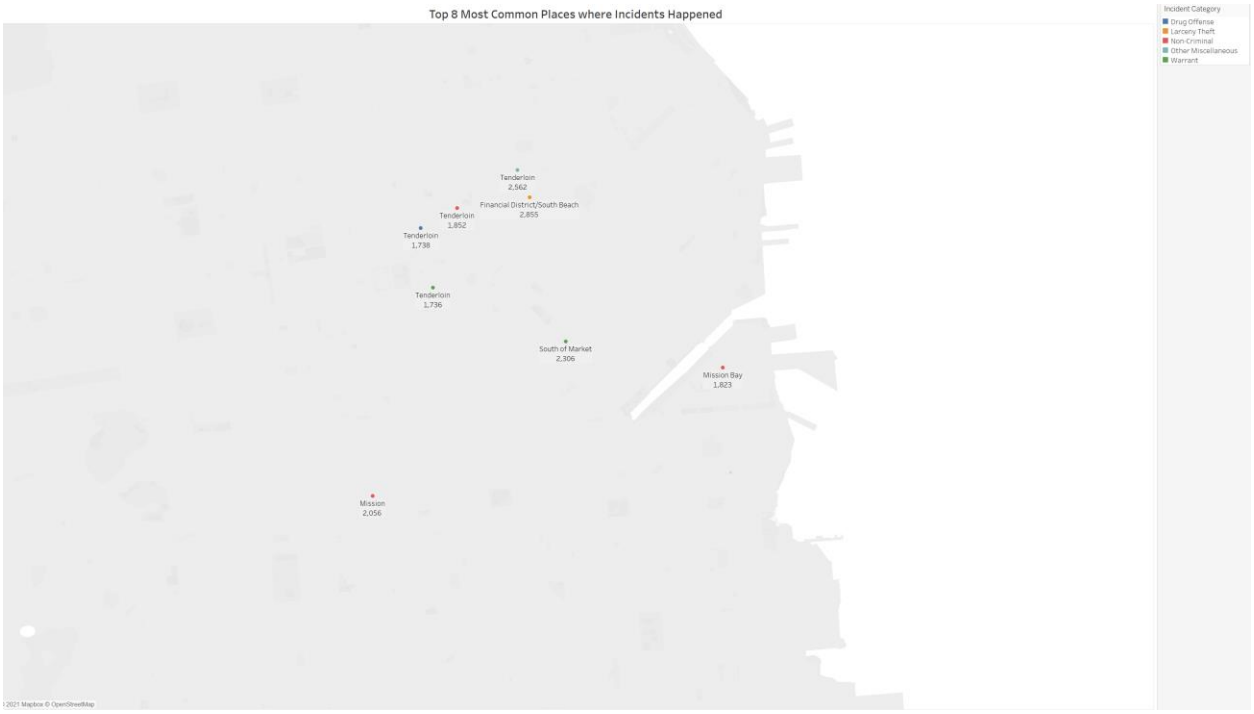
Share Phase:



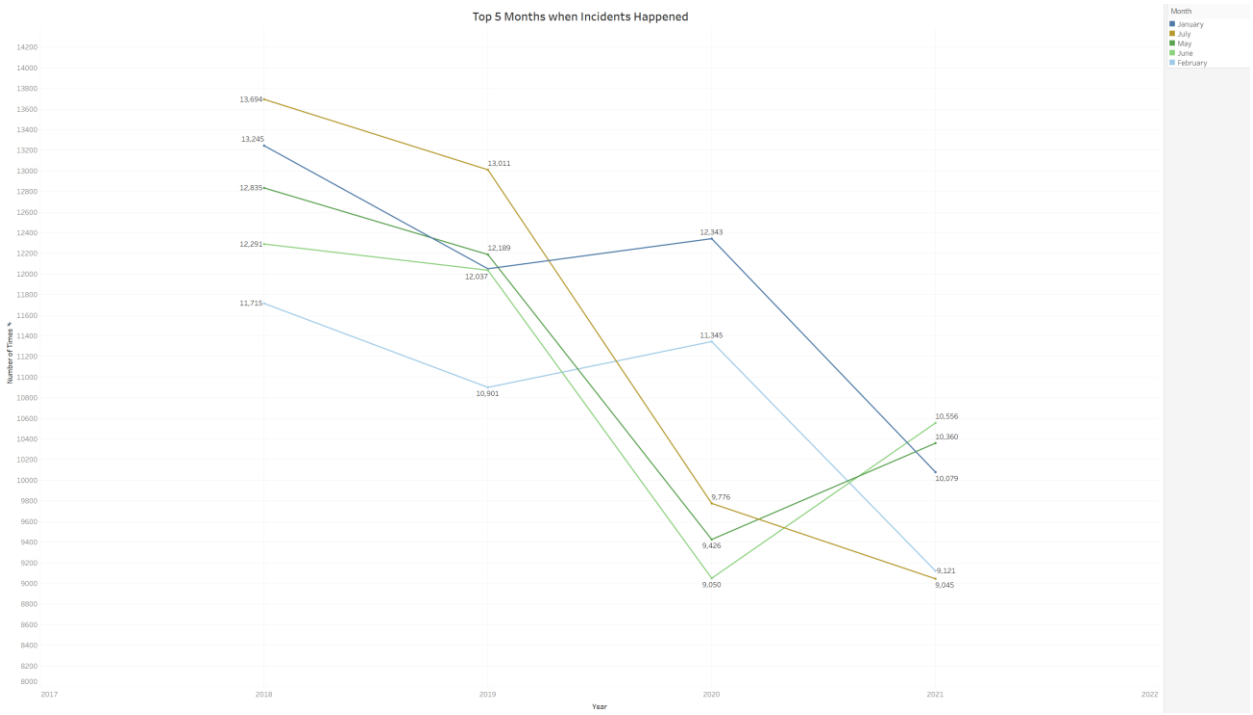
From the area graph above, the data shown the top 10 most occurred incidents types' trend from 2018-202. In most cases, the peak where the incidents happened were in 2019 and 2020. At 2021 all type of incidents were lower than the year of 2020 but still higher than 2018.



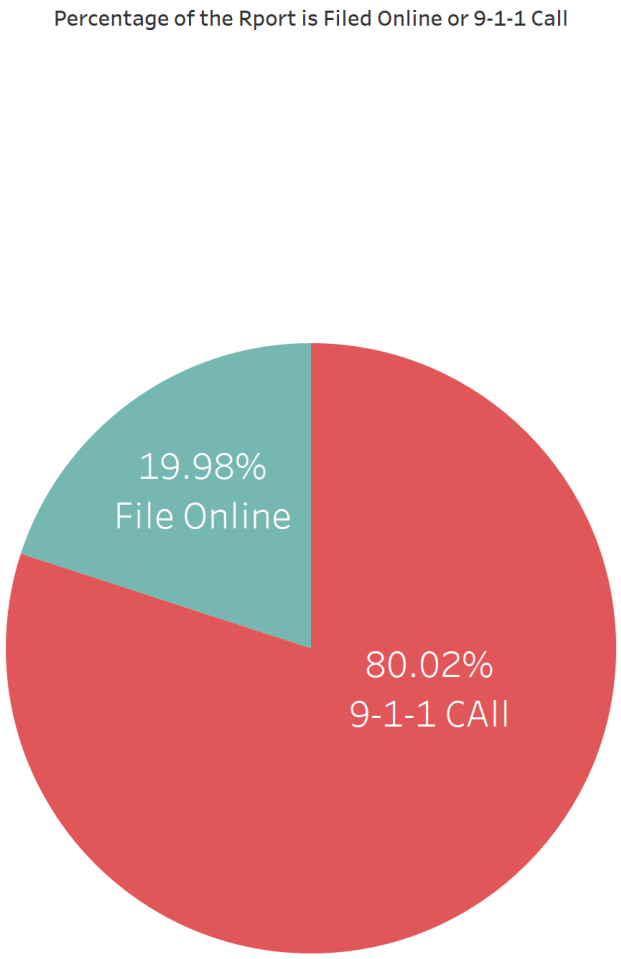
From the bar graph about, the data shown that Friday is the day where most incidents occurred for the past 4 years. Sunday is the least incidents occurred based on day of the week but the difference between rank 1 and rank 7 are actually not that much, only 9,913 incidents difference. In this case, we can assume that day of the week doesn't really have an effect on incidents occurred.



From the map above, the data shown the top rank neighborhoods where the incidents occurred the most. The neighborhoods are Financial District/South Beach, Tenderloin, South of Market Mission, and Mission Bay. The type of incidents that occurred the most at Financial District is larceny Theft. In this case increase patrol or police force at that neighborhood might increase the safety of people who lived there.



From the line graph above, the data shown the months when the incidents occurred the most through the year of 2018 to 2021. Because the data was only to July 2021, August, September, October, November, and December will not be in this graph. Back in 2018, July is the top month where the incidents occurred the most, but through out the year, the incidents occurred keep on decreasing in July and in 2021, it is the least incidents occurred month. As for the month when the incidents occurred the most in the year of 2021, June will be in first place. in the previous year, 2020, June has the lowest incident occurred rate but in 2021, June become the first.



From the pit chart, the data shown that 19.98% of incidents were reported online and 80.02% were reported through 9-1-1 call, which means almost 20% of the incidents were non-emergency and 80% if incidents were emergency.

Act Phase:

Recommendation:

1. Neighborhoods such as Financial District/South Beach, Tenderloin, South of Market Mission, and Mission Bay are the places where police should increase patrols or more police officer to decrease the rate of incidents and increase the safety of the neighborhoods.

2. Even though the incidents occurred rate are lower for most of the months for 2021 compared to 2018 but it is still an increase from 2020 so in this case, increased police officer is beneficial.
3. Most of the incident's reports are stilled not filed online which means most of them are reported through 9-1-1 calls, in this case, increase police officers will be beneficial to increase the safety of the city.

Data Collecting Recommendation

1. Record the time duration of the report and when the officer is onsite
2. Record the time duration when the incidents are solved or on going
3. Specify what type of warrants