

San Francisco Graffiti

W200 - Group 2 - Summer 2022

Collin Chee, Dan Nealon, Abel Ninan

Dataset

The dataset consists of reports of graffiti all over the San Francisco area from 2010-2022. The reports are compiled and stored by the city of San Francisco. The data is updated by the city on a daily basis. To keep our analysis consistent, we have decided to mark July 18, 2022 as the last date of analysis.

Link to Dataset

<https://data.sfgov.org/City-Infrastructure/Graffiti/vg6y-3pcr>

Dataset Structure

The dataset contains 731684 rows and 15 columns. (Variables in bold used for analysis)

- Caseld - Unique Id for each report
- **Opened** - Month/day/year time of day of case opened by the police
- **Closed** - Month/day/year time of day of case closed by the police
- Updated - Month/day/year time of day of case updated by the police
- Status - status of the case
- **Status Notes** - extra notes on the status
- **Responsible Agency** - which agency is responsible for dealing with the graffiti
- **Request Details** - whether graffiti is offensive or not offensive
- **Request Types** - where graffiti is physically (i.e. mailbox, pole, etc)
- Category - type of art of the graffiti
- **Address** - where the graffiti is reported to be
- **Neighborhood** - which district does the address belong to
- **Supervisor District** - San Francisco Supervisor District as defined in 'Supervisor Districts as of April 2012'
- **Point** - Latitude and Longitude of the area reported
- Source - how the request was received

Additional datasets

- San Francisco Police reports from 2018 - 2022:
<https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783>

Introduction

Graffiti is a drawing or painting on a wall or surface without permission and is visible by the public. This form of art is extremely popular and controversial as property owners consider it a form of vandalism. Although punishable by crime, many “artists” still display their artwork to spread gang-related activities, hate messages, and other messages. The city of San Francisco has kept a public database of all the reports of Graffiti within its territory since 2010. The dataset contains over 700k reports of graffiti in the San Francisco region with details about when it was reported and what type of graffiti is displayed on different surfaces. More detailed information about the data can be found [here](#).

By examining this dataset, our group hopes to answer the following questions:

1. Are there any monthly/seasonal trends in graffiti cases?
 - a. Any monthly/seasonal trends between a case is opened and closed?
2. How long does it take for opened graffiti cases to be closed?
3. Are the top 3 agencies in San Francisco specialized with respect to the type of graffiti they seek to mitigate?
4. Is there a certain kind of graffiti localized to a certain zipcode?
 - a. How varied are the types of graffiti that a given zipcode sees?
5. Which streets tend to contain the most reports of graffiti?
6. Do regular crime rate trends match the number of graffiti in certain sections of San Francisco?
 - a. We will be combining this SF incident dataset above that showcases incident reports since 2018 to compare graffiti and crime trends. We will join on time and in the zip code area.

Data Cleaning

Before we performed any analysis, we checked that the data contained all the correct information. Some information we needed to make sure of was:

- a. All locations in the dataset are actually in the San Francisco region.
 - i. In order for this to be valid, we first extracted the zip code in the Address column and filtered all rows based on their zip code. If the zip code was not in the SF region or was not specified in the Address column, we dropped those rows immediately.
- b. All reports of graffiti represent a unique graffiti on a surface
 - i. Luckily, the Status Notes column would comment on if a certain graffiti was a duplicate case. If the status notes contained the word ‘duplicate’,

we dropped them immediately and the original report of a certain piece of graffiti would be retained.

- c. Incorrect reporting on graffiti case times. We found that the time recorded between when the case opened and closed was inconsistent, with that the time of closing occurred at the same time (5 cases) or prior to when the case was even opened (3684).

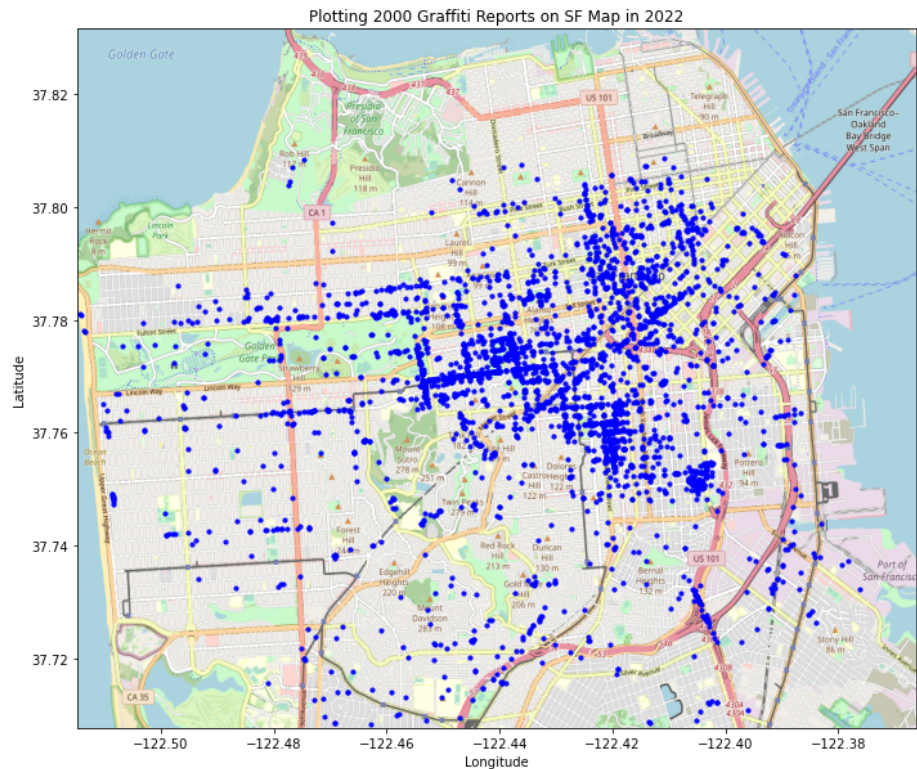
Time to Close a Case	Number of Occurances
Less than 0 seconds	3684
More than 0 seconds	650346
0 seconds	5

These incidents were removed from our analysis as we could not conclude if it was a bookkeeping error and they were just flipped or if there was a data error.

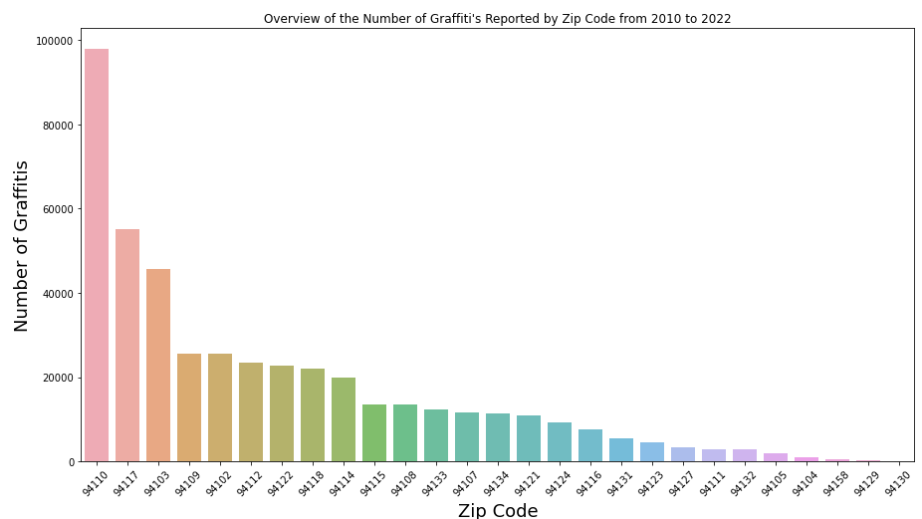
Initial Exploration

After doing a sanity check on the data, we did some initial exploration on the data set.

Map of graffiti reports in 2022: There were around 27,000 reports of graffiti in 2022 alone. The map here only shows 2000 of the reports for readability purposes. We can see that most of the reported graffiti are in central SF (Mission District) and reports are usually directly on the streets of SF.



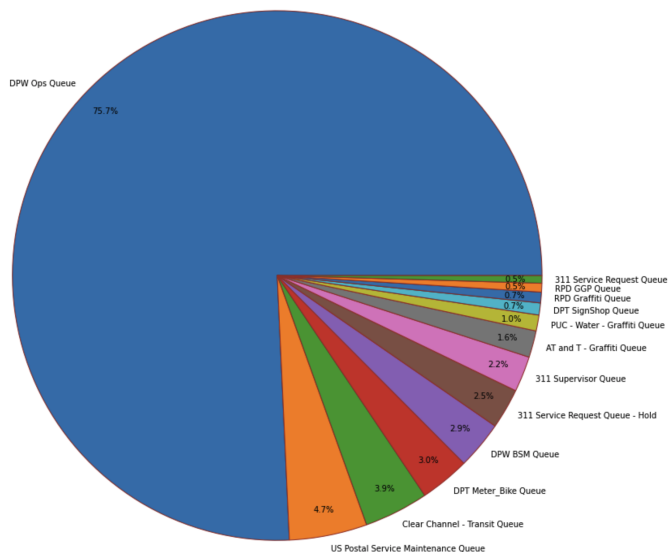
The following bar graph confirms that the majority of the graffiti reports are located in the Mission District as zip code 94110 is basically all of the Mission District. Further analysis was conducted and solely focused on this area code.



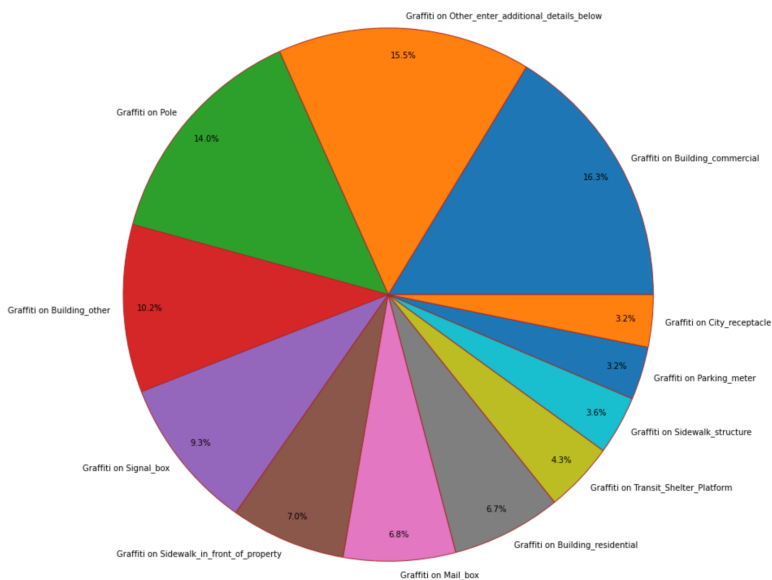
Graphs depicting the major agencies that address graffiti and the types of graffiti they come across: DPW Ops Queue addresses an overwhelming majority (~75%) of graffiti incidents in SF with a mix of other agencies addressing the remaining 25%. The types of graffiti

that are seen and cataloged are more varied and indicate that graffiti is found on many different objects likely located on any given street in SF.

Agencies Responsible for Addressing Graffiti Incidents

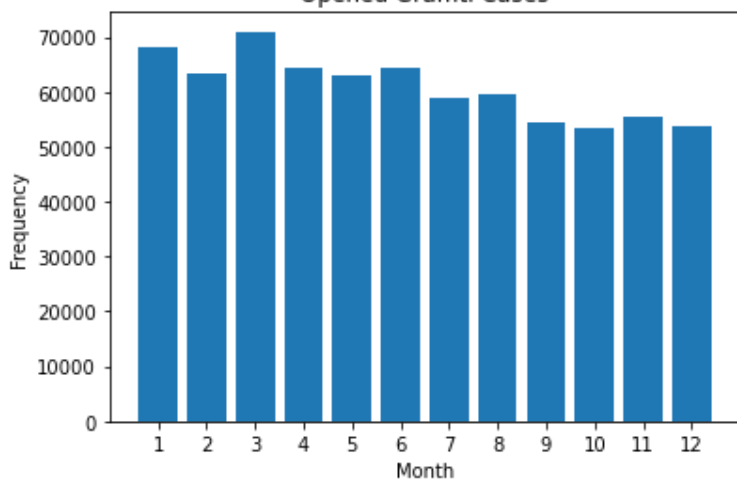


Types of Graffiti

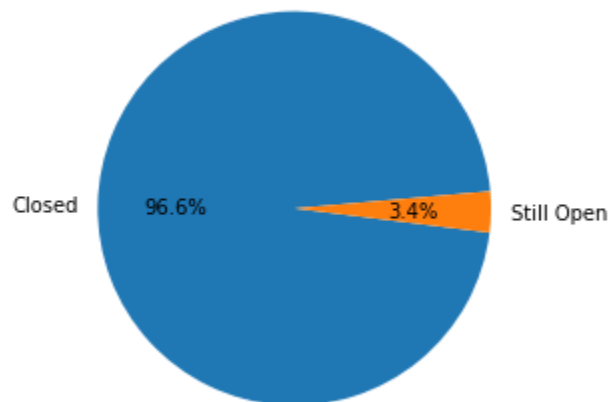


Graffiti cases by month: Hypothesized that more graffiti incidents when weather is nicer, maybe summer when kids are not in school. Do not really see this. Cases peak in the initial months of the year and follow a decreasing trend over the subsequent months. Could be something more going on, maybe more to learn about graffiti tendencies.

Opened Graffiti Cases



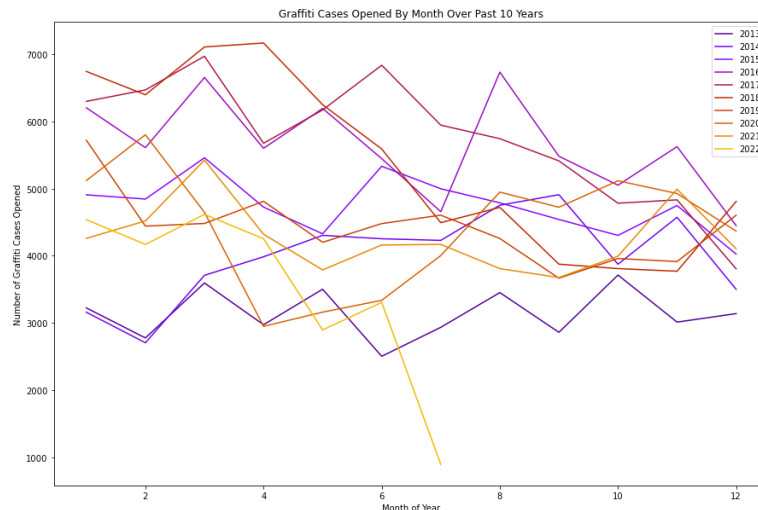
Percent of Graffiti Cases that are Closed



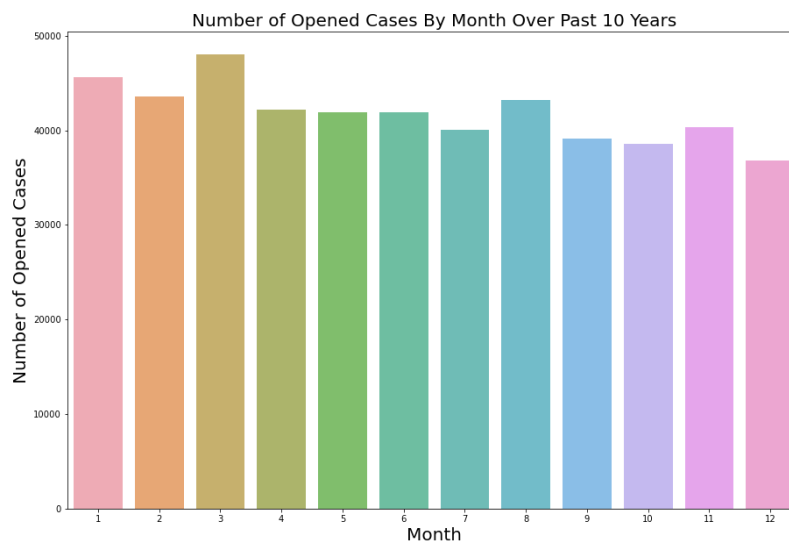
Questions

Are there any monthly/seasonal trends with graffiti cases being opened?

From our initial analysis, we saw that graffiti cases may slowly decrease month over month from January through December, however needed further investigation. Firstly, we broke out the opened cases by month and year.



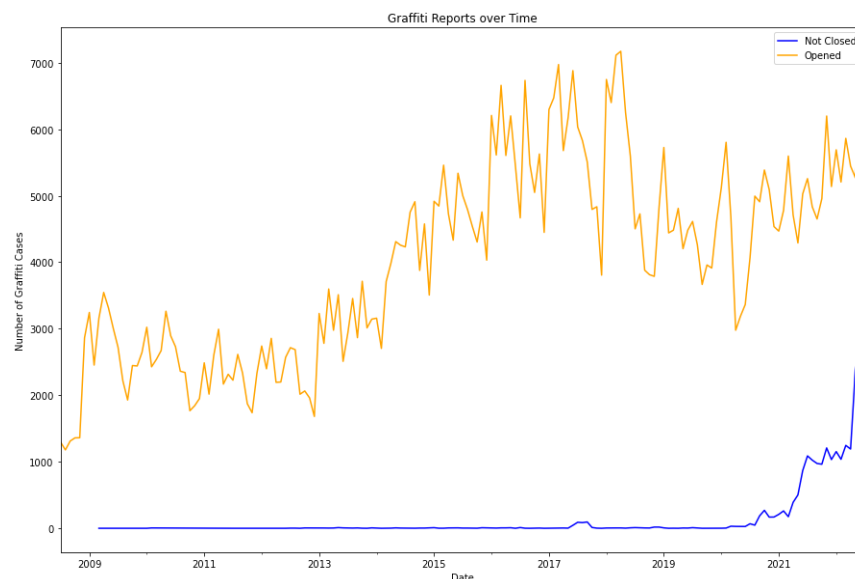
From this, with lines representing years and the lighter lines corresponding to more recent years, we do not see any distinct cyclical patterns suggesting common yearly seasonality. There does appear to be a slight spike in March and a decrease in December for most of the years. We do see the 2022 series dropping off in July, which is due to that being the current month and thus where data has been collected up to. We remove this series and then aggregate by month to confirm this finding.



We do see an uptick in cases in March and a decrease in December. There is a bump in August which corresponds to the spike experience in 2016 but would not be considered seasonality.

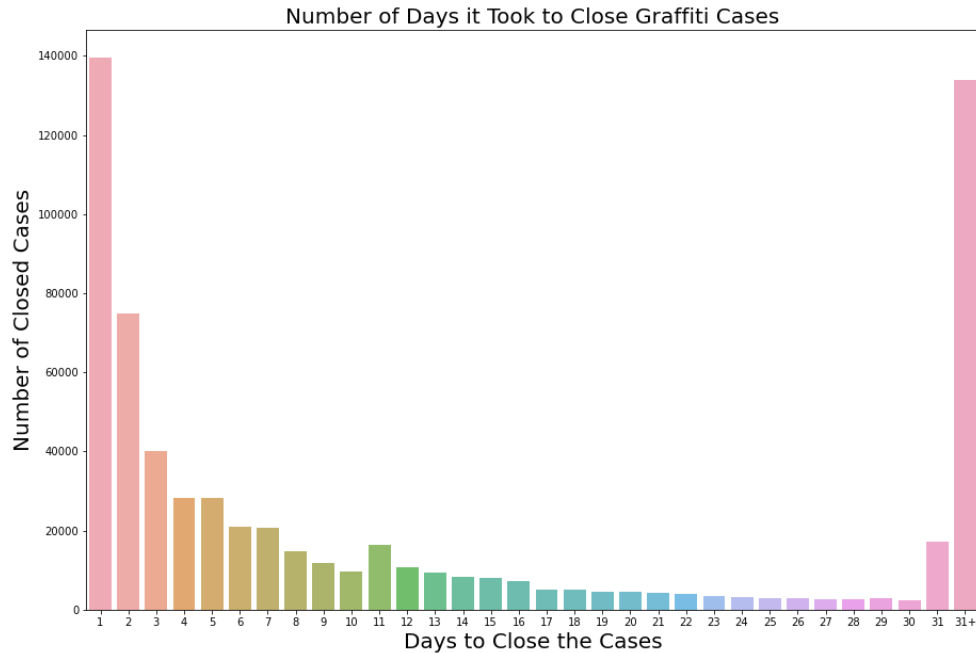
How long does it take to close graffiti cases?

From our initial analysis we saw that over 96.6% of opened graffiti cases were eventually closed, but we wanted to dig into this more. We first started by understanding the closed cases over time, and saw that the 96.6% is actually not the true 'closure rate', or percentage of cases that will be closed. The graffiti dataset we are using is constantly being updated with new graffiti cases and there is likely a lag time between when a case is opened to when it can be closed, the below plot of graffiti reports and their status shows us this behavior.

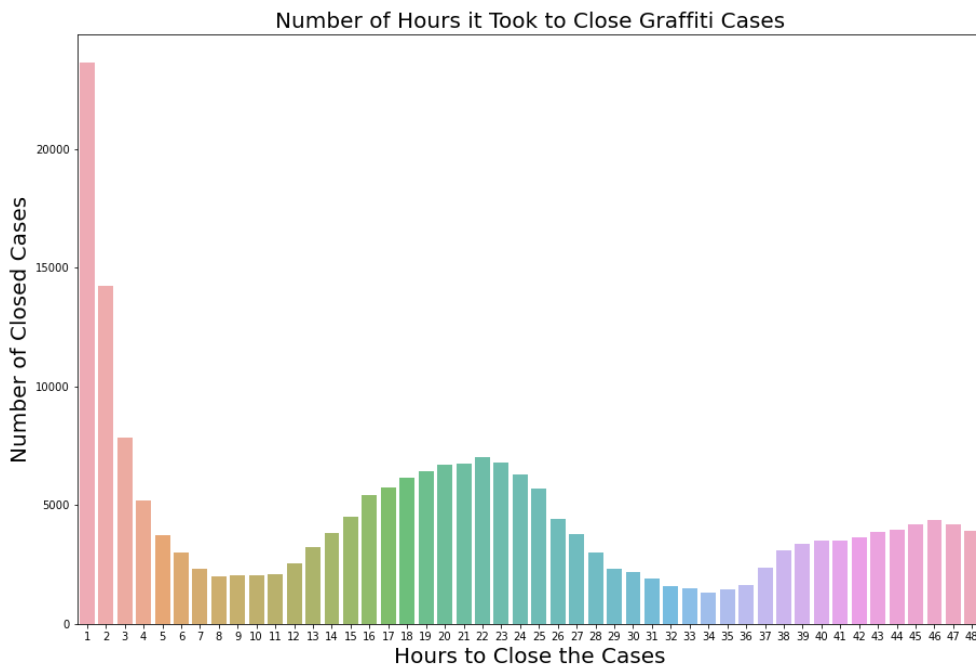


From this we can see that as 2021 progresses and we enter into 2022, the number of opened cases does not show any drastic changes, however the number of 'not closed' cases does suggest that these cases may end up being closed, however not enough time has passed yet.

This then led us to examining the time between the opening and closing of cases. At this point, we only wanted to examine cases that have been closed so the 3.4% of cases not closed were removed. Additionally, from the data cleaning section, we found that there were some closed dates that were before (or exactly the same as, down to the second) the opened dates, which did not make logical sense and were excluded for this section of the analysis. With the remaining data, we calculated the difference in time between the open and close dates of the cases and found the following.



We see that the first few days are the most frequent length of closure, however we do see a long tail and thus bucketed after one month for visualization purposes. Interestingly, the spike at the 31 day mark may suggest that there could be some rounding going on with how they track time between the opening and closing of cases and 'a month' being a common and acceptable standard to use. Within this, we zoomed in to the first two days to see the time to close in the most frequent lengths.

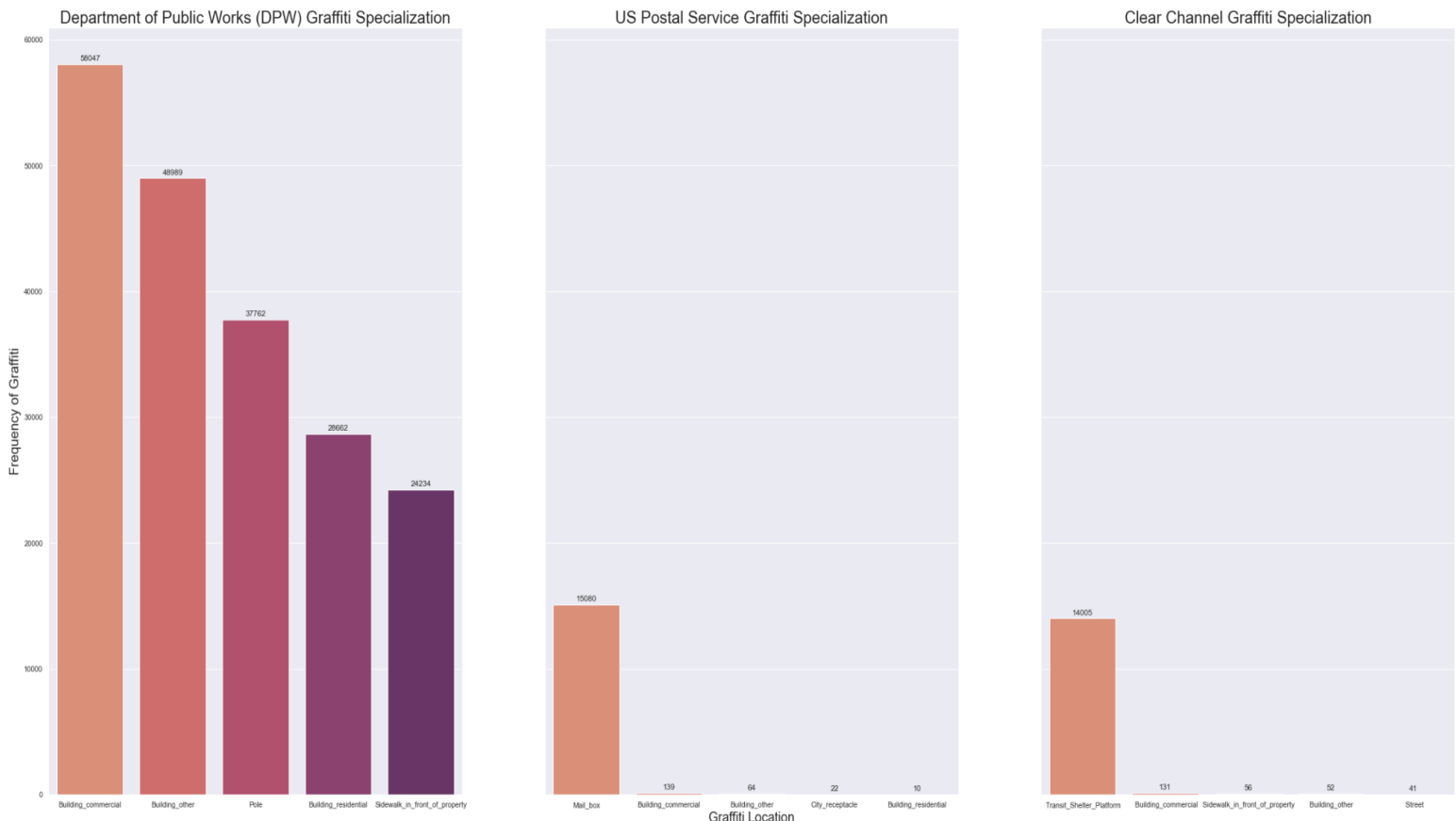


From this we see a distribution with an initial peak telling us cases are frequently closed very quickly but we also see another peak approximately 22 hours after the case was opened. This is likely linked to the work day, with the cases not being worked on from say 5 PM until 8 AM the next day, and then being closed the following work day. This behavior intuitively would be more common later in the day, as there is less time to work on the case before the end of the day, but being picked up at the beginning of their following work day, roughly 17 hours later. This aligns with the increase we see causing the peak at 22 hours. There is a similar pattern occurring around 46 hours with a local minimum at 34 hours, which agrees with this logic for when employees would and wouldn't be working to close cases.

The long tail behavior is shown in the mean closure time, which is 58 days while the median is only 5.6 days.

Are the top 3 agencies in San Francisco specialized in the type of graffiti they seek to mitigate?

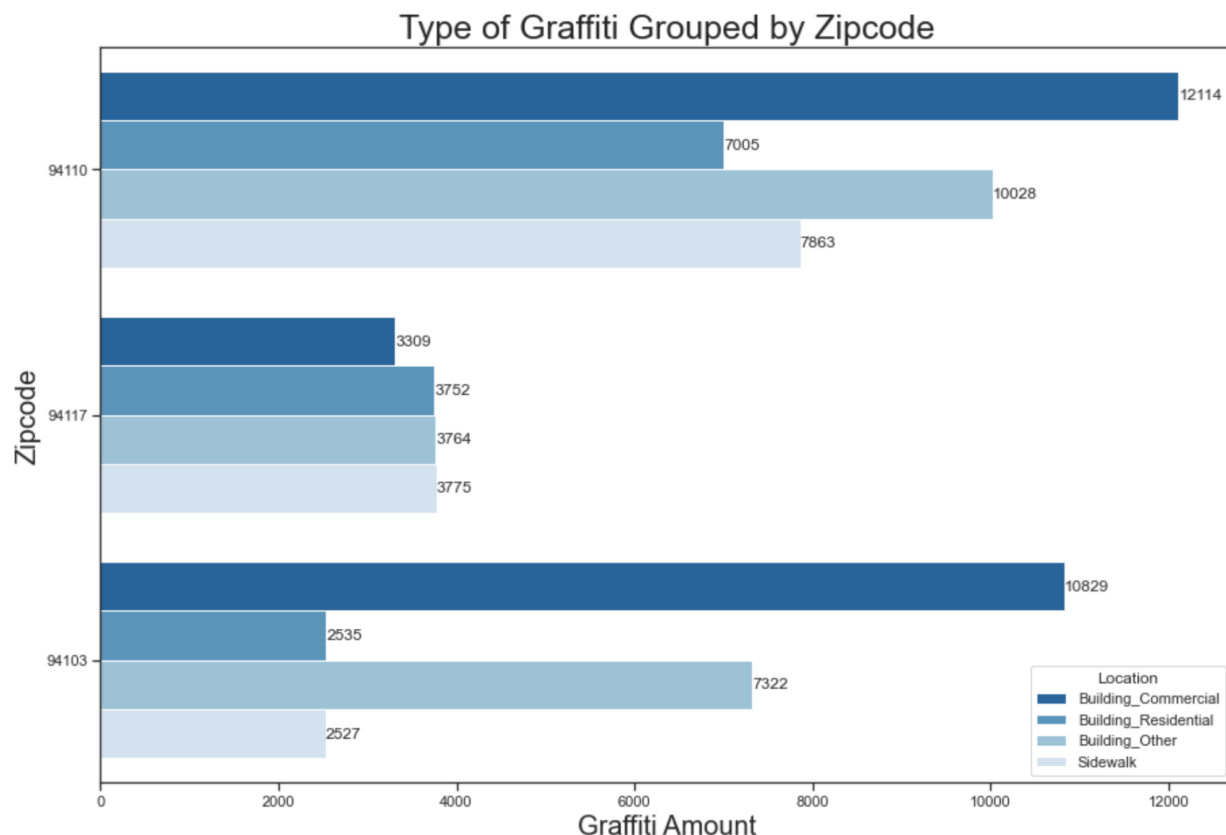
From our initial look into the dataset, we were able to see that there are 13 agencies in the city of San Francisco that are called upon to deal with graffiti incidents. However, most of the 13 only deal with relatively few instances when compared to their counterparts. In order to determine a concise answer to this question while also accounting for a majority of graffiti incidents in the city, we focused on the top 3 agencies listed in the database. The top 3 agencies, DPW Ops, US Postal Service, and Clear Channel, were responsible for addressing approximately 85% of the cases listed in the database.



As seen here, DPW handles the vast majority of cases but they are not without help. They do, however, seem to lack a clear specialization as they work with a substantial amount of graffiti in a variety of locations. US Postal Service and Clear Channel are associated with the opposite scenario. They do not deal with graffiti in such high volumes but do in fact almost exclusively work on mailboxes and transit shelters respectively, which makes sense given the organizations' business objectives. They do also help out DPW in addressing graffiti outside their specialization but not frequently. Regardless, these agencies helping out DPW (preventing them from being overwhelmed) is likely a major factor in why almost 97% of graffiti cases in the database are closed.

Is there a certain kind of graffiti localized to a certain zipcode? How varied are the types of graffiti that a zipcode sees?

For this question/analysis, we focused on three specific zipcodes, 94110, 94117, and 94103, as they contributed substantially to the total number of graffiti incidents reported in the database. We found that these zipcodes showed high levels of variety with respect to the unique kinds of graffiti present within its borders. However, there were four types of graffiti that were commonly found to be listed amongst the top 5 types of graffiti at all 3 of these zip codes. They were graffiti found on: Building_commerical, Building_residential, Building_other, and Sidewalks.

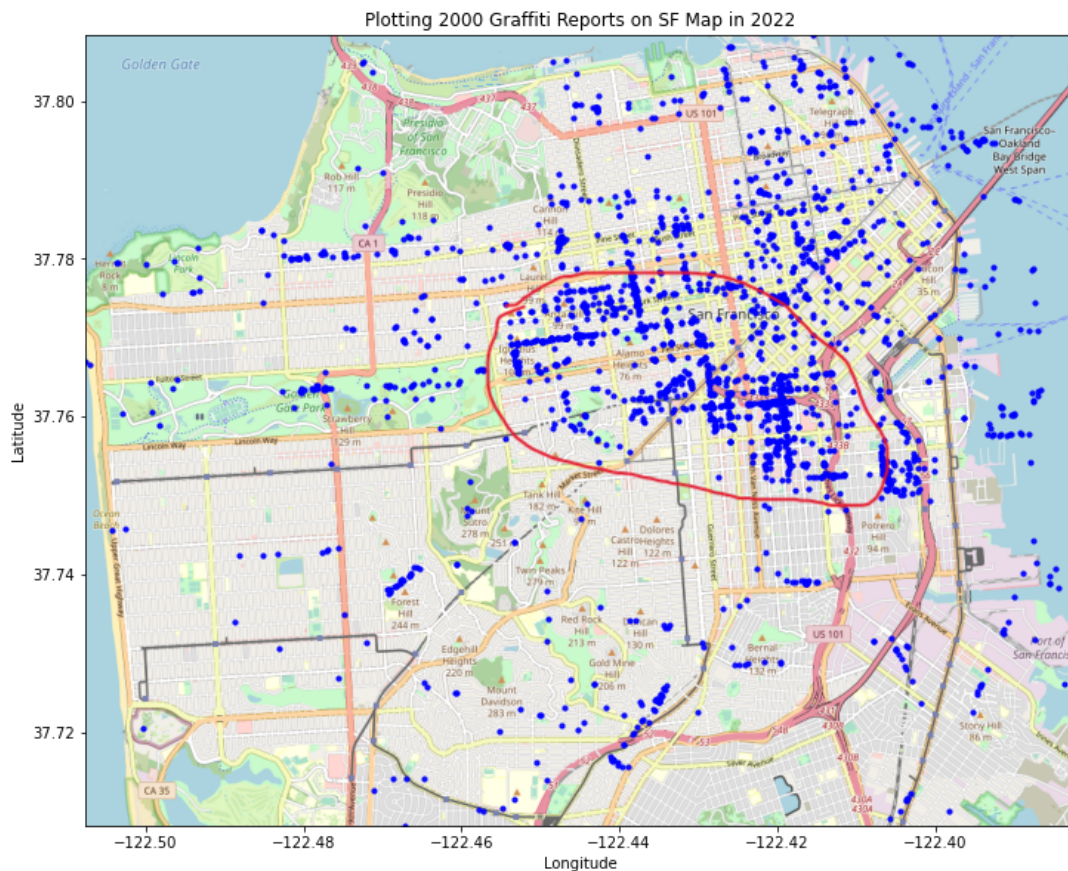


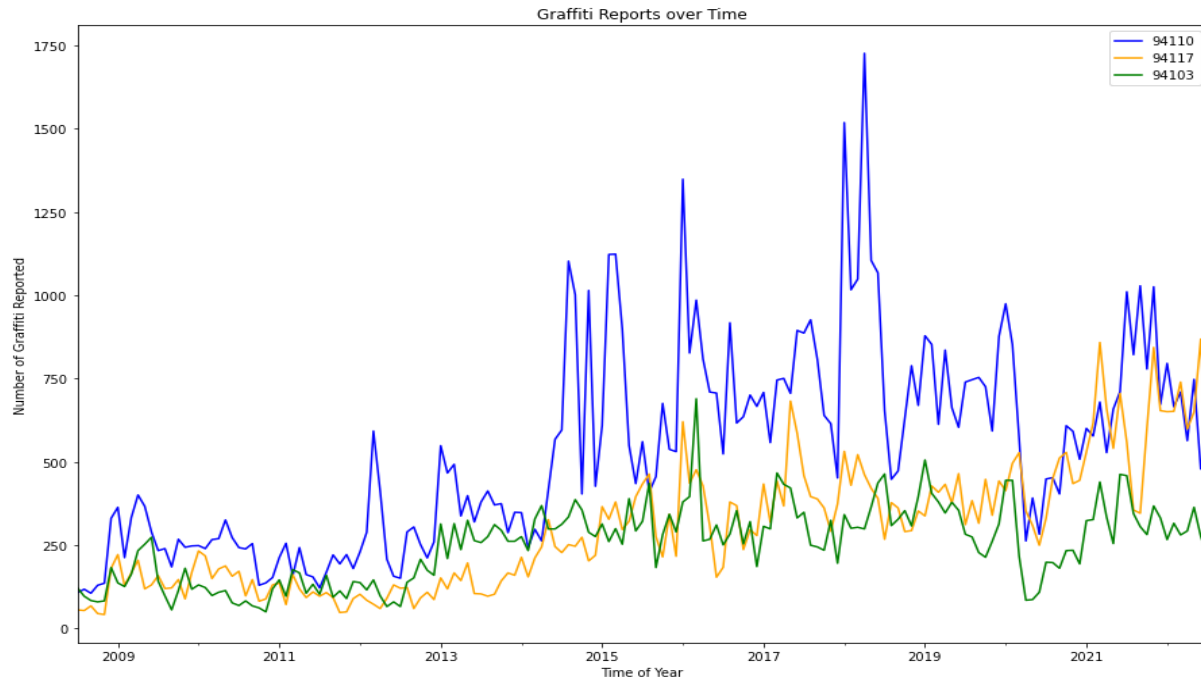
Zipcodes 94110 and 94113 show similar patterns regarding where graffiti is most-commonly and second-most-commonly found. Zipcode 94117 stands out considering that they have all four locations displaying similar frequencies of graffiti amount. These 4 types of graffiti at all 3 zipcodes coincide with 4 out of the top 5 types of graffiti that DPW (the leading agency in the database) addresses.

Streets that Contain the Most Graffiti

Based on the initial exploration of the zip codes with the most graffiti reported, we will look at zip codes 94110, 94117, and 94103 since they had the top 3 highest number of graffiti reports. Perhaps the most popular streets contain the most offensive graffiti. Here is a reminder of the area we will be specifically exploring.

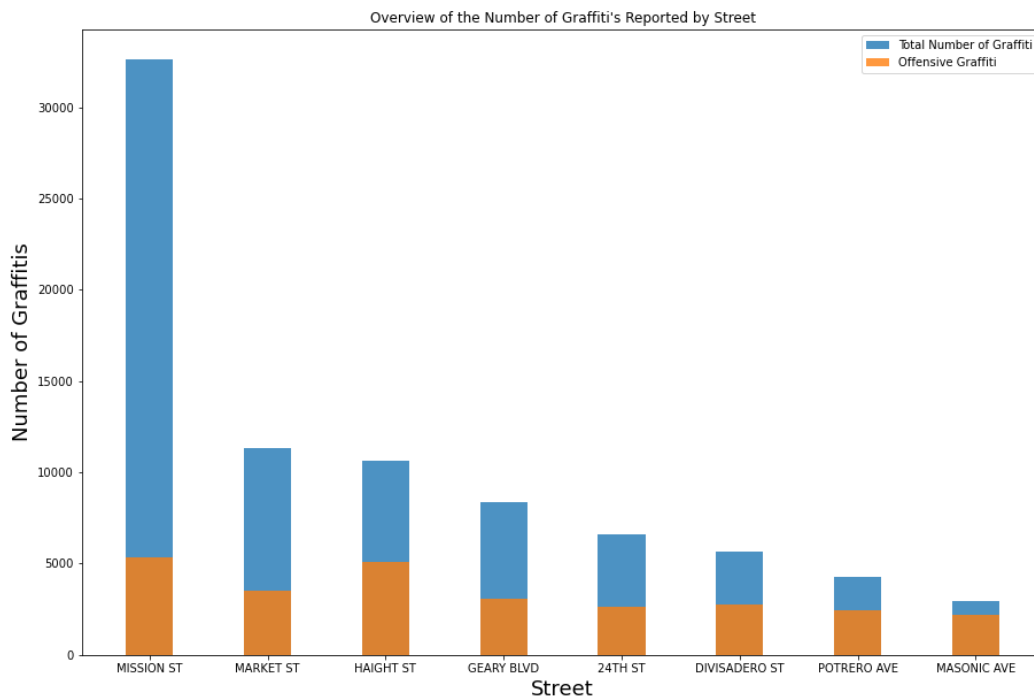
Although we see that zip code 94110 is the hot zone for graffiti, has it always been the most popular artwork area or has it recently received the most attention?





As we can see from the timeline above, 94117 at times has seen more graffiti reports in 2022. However, 94110 in the past decade has comfortably been the most popular area for graffiti reports.

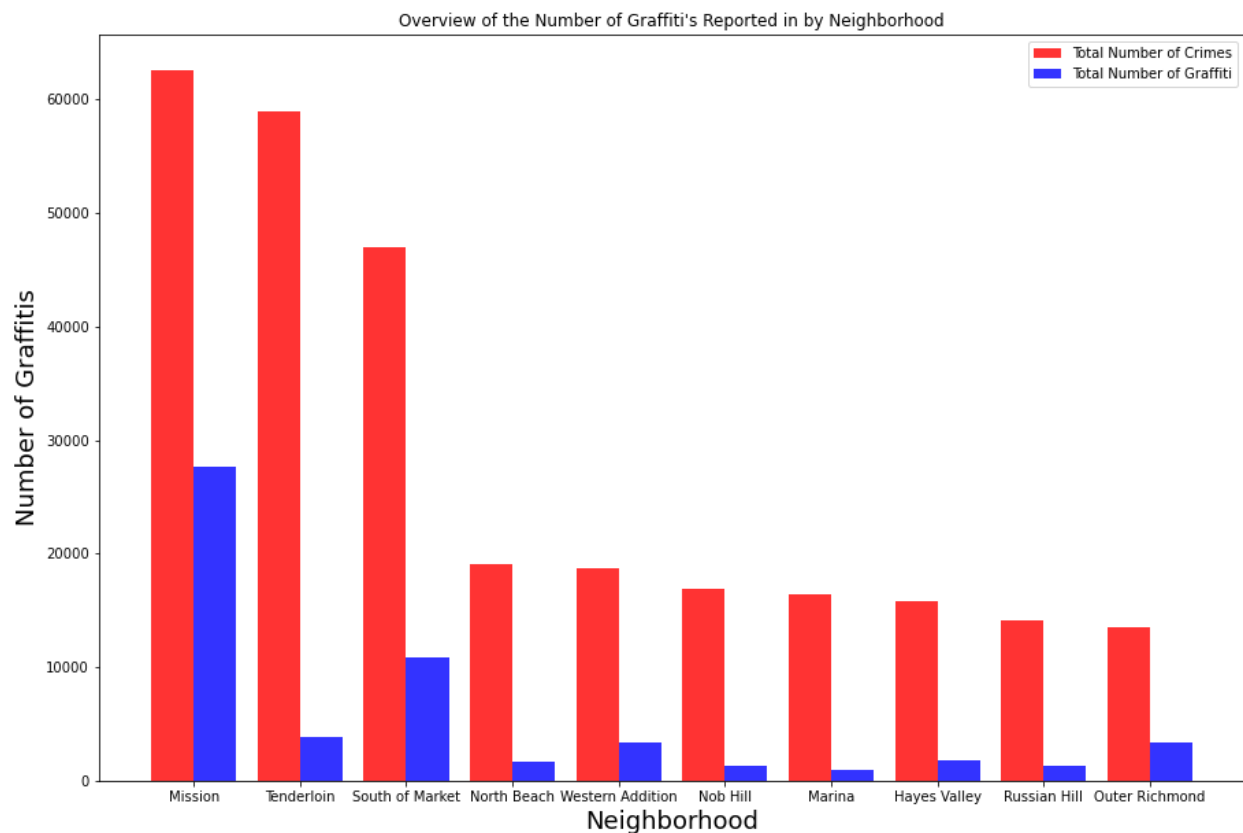
Let us look at within these three zip codes, which streets contain the most graffiti, and are they offensive?



As illustrated in the bar graph above, Mission Street is the heart for most of the graffiti. No surprises that Mission Street is also part of the 94110 district. However, what is interesting about the graph above is that although Mission, Market, and Haight street receive a lot of attention in the graffiti world, the city of San Francisco does not deem most of the artwork as offensive . Other streets such as Divisadero, Masonic, and Potrero street appear the most notorious for displaying offensive graffiti. Of course, graffiti in of itself is still considered vandalism, but perhaps people in certain areas like to express less provocative images.

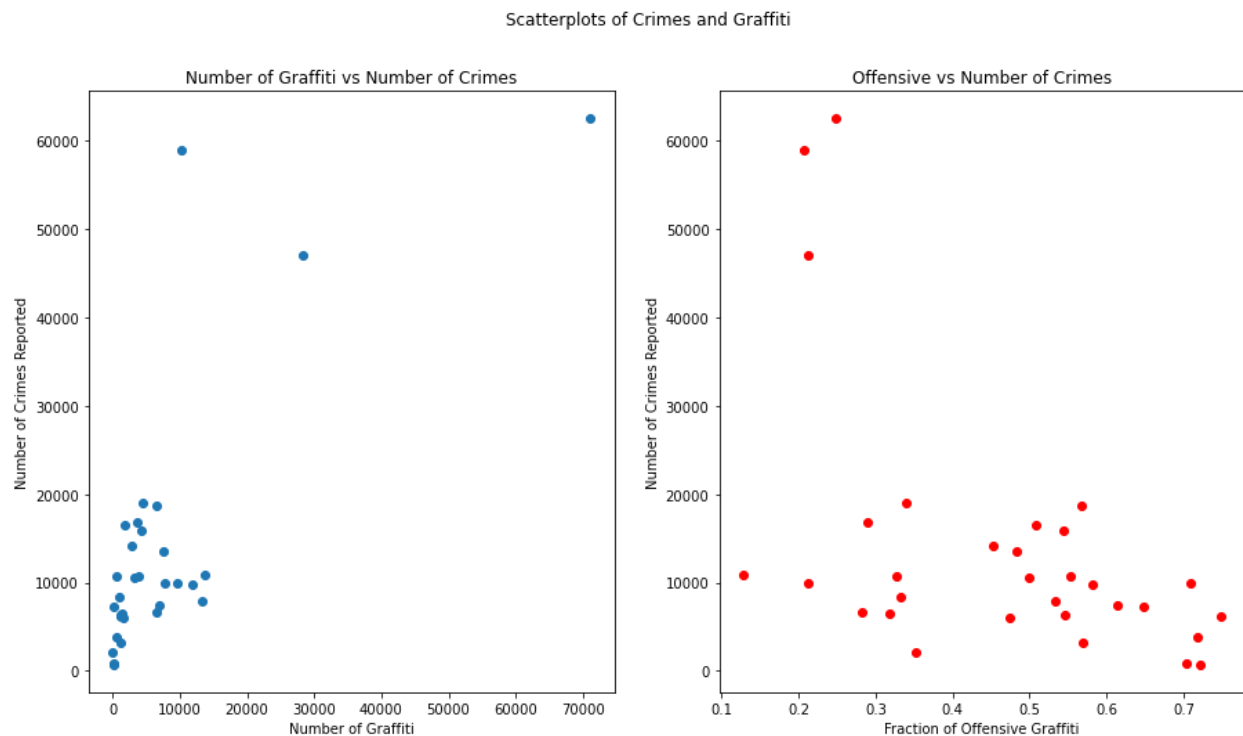
Comparing Number of Crimes and Graffiti

Because graffiti is viewed as something associated with gangs and disturbing activities, we will explore if this preconceived notion is actually true. Unfortunately, the majority of the crimes reported did not have a direct location of where the crime was committed. Therefore, this analysis will consist of comparing graffiti and crimes grouped by their neighborhoods.



The bar graph seems to vary a lot with whether there is a correlation between the number of crimes and the total number of graffiti. An interesting takeaway is that the

Mission Neighborhood is the only neighborhood in the top 10 with more graffiti than crimes reported.



In this side by side plot, we can see on the left scatterplot that there could be a positive correlation with the number of graffiti reported and the number of crimes reported. However, perhaps the type of graffiti does not really matter. Even areas with majority offensive graffiti, the crime rate is actually lower.

Conclusion

Our analysis of the San Francisco graffiti helped us answer most of our questions. The data set allowed us to understand trends with graffiti over time and showcased areas with the most graffiti. We also learned about when and where graffiti cases are opened the most and how the city of San Francisco responds to these reports.

The city of San Francisco has done extensive work to address many of the graffiti and as we can see above it may be related to reducing criminal activity in general. In order to draw stronger conclusions about connecting graffiti and criminal activity, there are many other labels to explore such as financial and environmental factors differences between San Francisco districts.

Some future research on this topic can include:

- Although we analyzed how the time to close cases vary, we do not know why

they vary. Research can look into the processes of addressing graffiti.

- Crime and graffiti have many motivations. What are they?
 - Survey actual artists to understand why people present their graffiti in certain areas
 - Certain areas have significantly more graffiti than other districts. Why is that?
- Potential impact of the city allocating funding/resources to the smaller organizations that address graffiti to see if they can bolster their numbers and contribute to addressing graffiti at a higher volume