Twitter User Sentiment and Crime in San Francisco

Atmika Sarukkai

UC Berkeley

Data Science Honors Thesis

May 2023

Abstract

This thesis aims to analyze the relationship between Twitter user sentiment on crime in San Francisco and actual reported crimes in the city. This work seeks to understand user perceptions of different kinds of crime and identify whether Twitter user perception may serve as a useful indicator of criminal activity. In conducting a time series analysis of Twitter negative sentiment intensity and the number of San Francisco Police Department (SFPD) crime incident reports for four different kinds of crime (person crime, property crime, drug crime, and minor offenses), we may compare similarities in variation and directionality of trends for each crime over the span of three months. The exploration of monthly trends reveals more granular similarities in trends between the two datasets. Word-vectorization of tweets and dimensionality reduction allows for an exploration of the associations between crime-related words and fear-related words. The analysis suggests that monthly trends in person crimes (violent crime and sex crime) and minor offenses for Twitter negative sentiment and SFPD incident reports share similarities. Our Word2Vec model suggests that person crimes and drug crimes may be more associated with fear sentiment than property crimes or minor offenses.

The public GitHub repository for this thesis can be found at:

https://github.com/atmika-sarukkai/Honors-Thesis

1. Introduction

According to a survey conducted by the SF Chronicle of San Francisco residents earlier this year, safety and crime is the most pressing issue facing the city, second only to homelessness. During the peak of Covid, the city observed increases in crime rates, specifically larceny and theft related crime, which disproportionately affected certain communities, such as Black and Asian American communities. This project focuses on public response to fluctuations in crime in San Francisco over the last decade in parallel to analyzing trends in reported crime in San Francisco. This comparison of trends in public sentiment and crime may expose certain adequacies or inadequacies of law enforcement to address fear-invoking crime in San Francisco. This project may also reveal if social media, specifically Twitter, is a useful tool in reflecting trends in crime, as well as whether Twitter sentiment can potentially be used as an additional indicator of crime in the city.

This paper attempts to answer two central questions: Does the intensity of Twitter negative sentiment reflect suspicious activity and crime in San Francisco? Are certain kinds of crime more closely associated with fear than others? An analysis of trends in Twitter negative user sentiment intensity over time as well as an analysis of trends in the number of San Francisco Police Department (SFPD) incident reports demonstrate that user sentiment does seem to reflect general trends in variability and directionality of crime incident reports over time for certain crime types. Additionally, a neural network Word2Vec model of Twitter text reveals that certain crimes are more proximate to fear-related sentiment.

Literature Review

In understanding the relationship between fear and crime, we analyze literature from previous studies. One study conducted on a sample of 1,047 random telephone numbers reveals that 4 out of 10 Americans are highly fearful that they will become victims of violent crime, specifically murder, rape, robbery, and assault. 4 out of 10 Americans also feel unsafe in their everyday environments due to a fear of crime, and this pervasive fear crosses all demographic boundaries (U.S. Department of Justice, 1980).

Additionally, previous studies have revealed that fluctuations observed in the number of crimes suffered do not necessarily lead to increases or decreases in the general concerns of crime. Instead, the public response and fear relate to interactions between individuals and media coverage, so violence and fear can not be understood by focusing exclusively on victims (Curiel, 2018). In addition, observations on an individual level are typically based on victimization surveys, which do not frequently track the fears of the same individuals over time. Therefore, analyzing user sentiment on Twitter may allow us to understand broader, complex social perceptions based on the interactions of individuals regarding different kinds of crime, and may help us to understand social consequences of crime and violence.

In a 2020 Gallup poll asking respondents how often people worry about various kinds of crime, 71% of individuals responded that they frequently or occasionally worried about having your personal, credit card or financial information stolen by computer hackers, 67% worried about being the victim of identity theft, and 40% worried about having their home burglarized. In comparison, only 17% were frequently or occasionally worried about being murdered (Reeves, 2020). Even as more violent forms of crime may elicit more intense fear in individuals, it is less violent and more common forms of crime that elicit the most frequent fear.

One paper explores crime and fear on social media, discussing how tweets containing crime-related words may not necessarily be associated with crime, and how even manual inspection may result in uncertainty regarding whether tweets are related to crime or not. Additionally, social media displays a bias towards violent crime as compared to other crimes. Social media posts also offer scarce city specific crime related data and therefore may be difficult to use for analyzing fear of crime at a city level (Curiel, 2020).

Another paper contributes novel approaches to classification of crime and non-crime tweets in India using text mining based approaches. Crime tweets are distinguished based on the presence of distinguishing crime related words, such as fraud, kidnapped, and murder (Ristea, Alina). For data pre-processing, sentence tokenization, hashtag expansion, stop words removal, acronym and slang treatment, punctuation removal, and TF-IDF conversion is performed (Shah, Parthvi). The authors also perform crime tweet classification, utilizing four different machine learning classifiers to identify which type of classifier is most suitable for detecting crime tweets. Ultimately the ZeroR model was most accurate in identifying crime-related tweets (Lal, 2020).

Data

Data Collection

Twitter Data

To investigate the relationship between Twitter User Sentiment and criminal incidents in San Francisco, I leverage data collected using the Twitter API and publicly available San Francisco Police Department (SFPD) data on historical incident reports. Twitter data was accessed through the Twitter Developer Portal, where users have access to tweets from the past week. Tweets were filtered based on the presence of #sf or #sanfrancisco within the tweet text, to identify if tweet content pertained to the city of San Francisco. I accumulate weekly tweet data from January 31st to April 18th, containing ~ 66,500 instances of tweets.

SFPD Incident Report Data

For data on crime in San Francisco, I use publicly available SFPD data containing ~50 features describing criminal incident reports from January 31st to April 18th. The relevant information for this work include the type of crime that was committed from 45 different crime categories (ex: Larceny/Theft, Fraud, Vandalism, etc.), and the date that the crime was committed.

Data Processing

Extracting Crime-related Tweets

To extract relevant crime tweets from our database of San Francisco related tweets, I utilized organized Crime Classification Codes created by Mac Arthur (Arthur, 1999). These codes listed the characteristics of seven different crime categories, including violent, potentially violent, other crimes against person, sex, property, drug, and minor offenses. For the purposes of our study, we combine the violent, potentially violent, other crimes against person, and sex categories into a “person” category.

An initial step in extracting crime related tweets from the Twitter database required creating a list of crime related words. These crime words were obtained from a curated set of keywords that are associated with various forms of crime, derived from a dataframe on FigShare, an open access repository. The dataframe identified keywords for various types of crime categories including sexual crime, violent crime, organized crime, property crime, robbery of person, burglary, car theft, shoplifting, murder, rape, kidnap, vandalism, and guns. The database contained words in both Spanish and English, with a majority of words in Spanish. I used the google translate library to translate words from Spanish to English. I also modified the crime categories assigned to words within the FigShare database so each word was assigned one of the four crime categories (person, property, drug, or minor offenses). This database contained both crime related words in English and Spanish, so the English words were included in the final crime words database. It was important to include many possible crime-related words to ensure that most crime-related tweets in our database were included in our resulting crime database. As a result, general violent words, synonyms of crime related words, and words present in the MacArthur Crime Classification Codes were manually added to our crime words list with their associated crime label. It is important to recognize that the crime words ultimately selected and their corresponding crime types were subjectively decided, and results of this study may change depending on what words are included and how they are labeled.

To filter crime-related tweets from our database, tweets that contained at least one of the words in the crime words database were selected. The tweet would be assigned the same crime label as that of the crime word present in the tweet. If a tweet contained two or more unique words in the crime word database, the most common crime type was chosen as the tweet’s crime label. In the case that the tweet contained the same number of words in two or more different crime categories, the crime category of the crime word that appeared first in the tweet would be chosen to assign to the tweet. The final crime tweet database contained 3,984 instances.

SFPD Incident Data

The SFPD Incident Report data contained 47 crime categories. Incidents within the four crime categories were filtered from the database. Crime types were combined to include within each crime category.

Person Crime	Property Crime	Drug Crime	Minor Offenses
Robbery Offenses Against The Family And Children Assault Weapons Carrying Etc Weapons Offense Human Trafficking, Commercial Sex Acts Suicide Arson Homicide Sex Offense Weapons Offense Rape	Motor Vehicle Theft Malicious Mischief Larceny Theft Burglary Stolen Property Recovered Vehicle Fraud Vandalism Forgery And Counterfeiting Vehicle Impounded Lost Property Embezzlement	Drug Offense Drug Violation	Traffic Violation Arrest Other Miscellaneous Traffic Collision Suspicious Occ Warrant Disorderly Conduct Prostitution Case Closure Liquor Laws Suspicious Gambling

Table 1: Table of crime types included in SFPD incident report data, and corresponding crime label (person, property, drug, minor offense)

EDA

San Francisco Tweets

To address the first research question of whether negative Twitter user sentiment is reflective of crime in San Francisco, a time series analysis is performed to identify key similarities in variance and slope of trends within the two data sets.

To identify tweets with negative sentiment, the TextBlob library for processing textual data was utilized. A score between -1 and 1 is assigned to each tweet, where values greater than 0 indicate positive sentiment, and values less than 0 represent negative sentiment. Tweets assigned a value of 0 represent neutral sentiment.

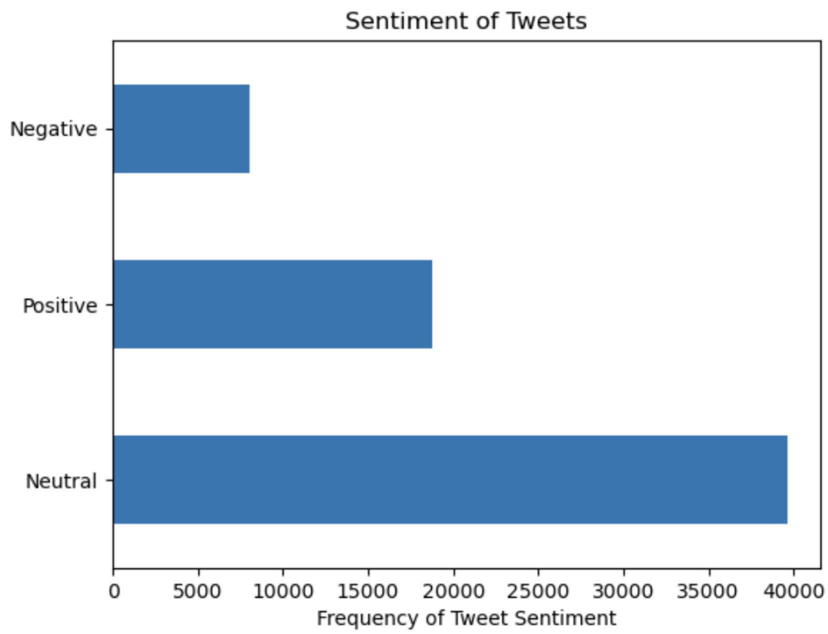

Figure 1: Count of Tweets with #sf or #sanfrancisco with negative, positive, and neutral sentiment. Most tweets have neutral sentiment, and the smallest number of tweets have negative sentiment.

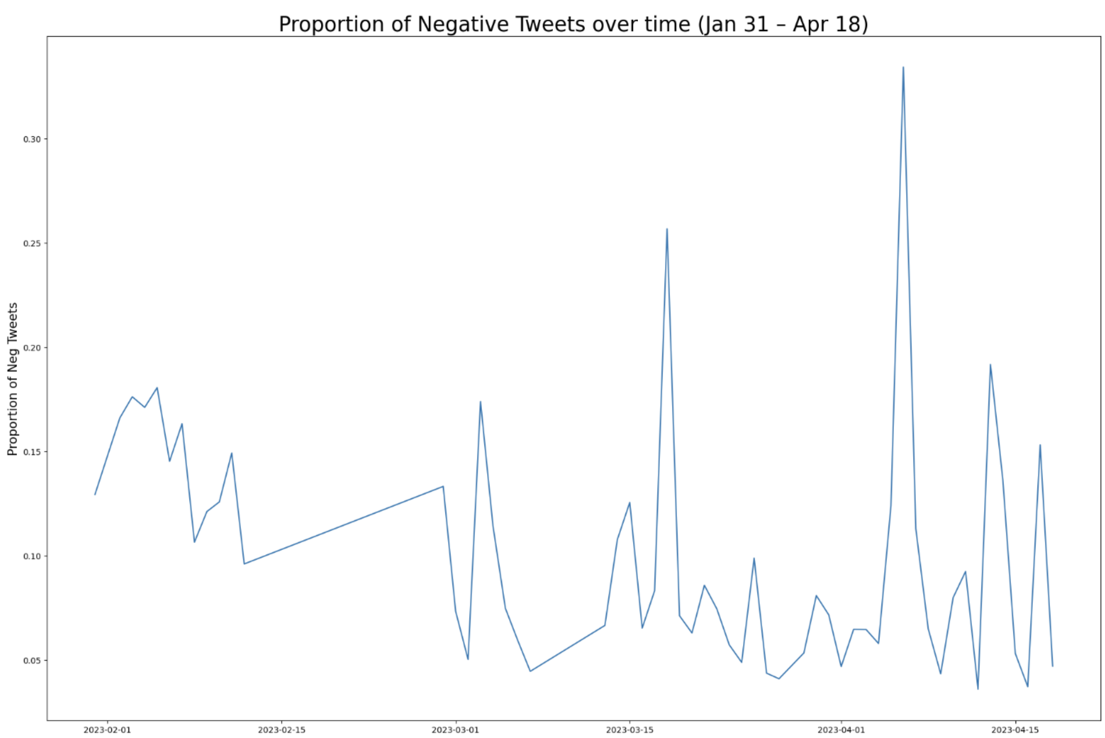

Figure 2: Line plot of the proportion of negative tweets daily from January 31st to April 18th. We may observe a general negative trend over time, and some sharp increases in the proportion of negative tweets in the end of March and beginning of April.

Methodology

Analysis of Overall Trends

San Francisco Crime Tweets

To analyze the overall trends of crime-related tweets, the intensity of negative sentiment for a given tweet is calculated using the VADER (Valence Aware Dictionary and sEntiment Reasoner) NLTK module's SentimentAnalyzer. The sentiment score of a tweet is calculated by combining the sentiment scores of each VADER-dictionary-listed word in the sentence. For each tweet, a polarity score is calculated, which calculates the probability that a tweet is negative, positive, or neutral.

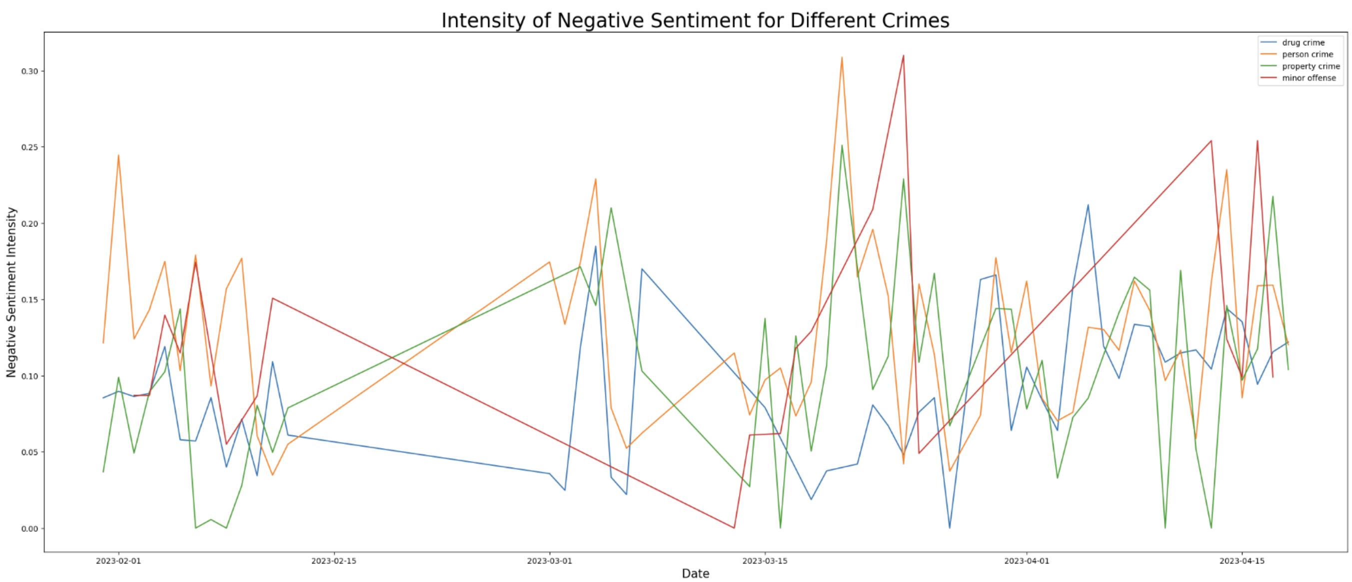

Figure 3: Line plot of the average negative sentiment intensity of tweets daily for each of the four different crime types from January 31st to April 18th

SFPD Incident Reports

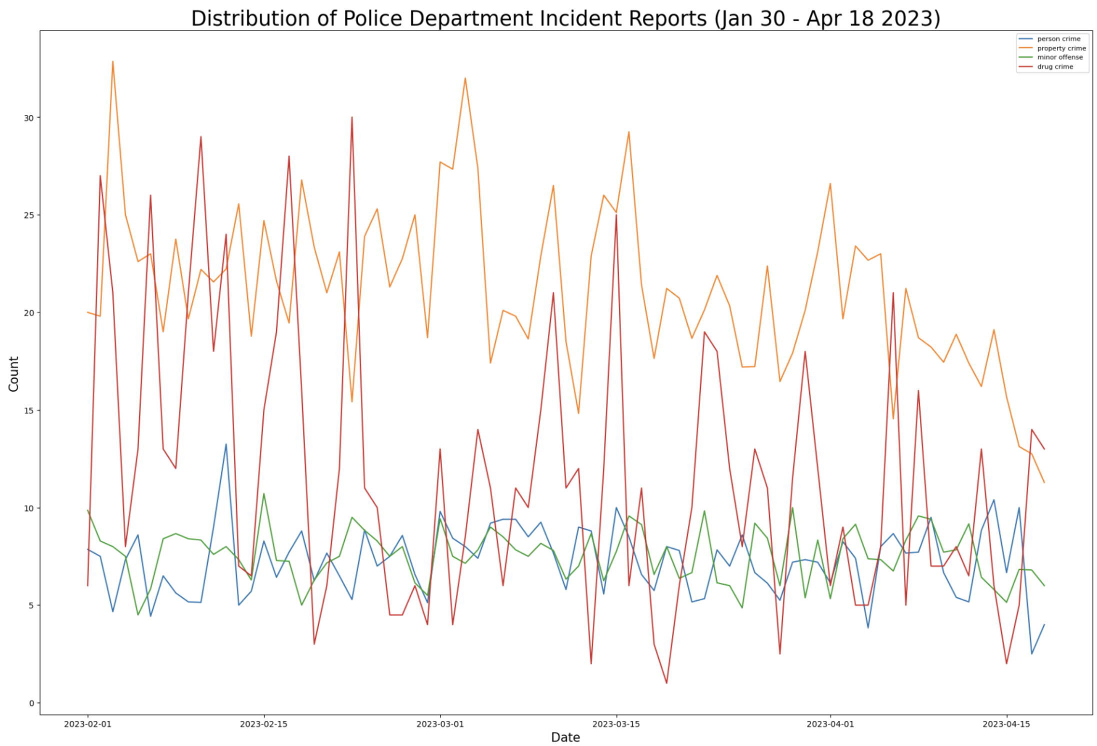

Figure 4: Line plot of the daily number of incident reports for each of the four different crime types from January 30th to April 18th

To analyze the similarities between trends of Twitter negative sentiment over time and the number of SFPD incident reports over time, two statistics are considered. To examine variability and fluctuation of the trends, a Mean Absolute Deviation (MAD) is calculated:

$$\frac{1}{n}\sum_{i=1}^{n}|x_i - m(X)|$$

Where m(X) is the average value of the dataset, *n* is the number of observations, x_i is the data value. Calculating a MAD for both datasets allows comparison between the datasets, and does not require that the data is normally distributed, unlike a statistic for variance such as standard deviation.

To quantify directionality of the trends, a line of best fit is created for each crime in both Twitter data and SFPD data. To create a line of best fit, linear regression models are created for each of the four crimes, fitting negative sentiment score for Twitter data and number of incident reports for SFPD data using the sklearn library's LinearRegression method. The coefficients were then extracted from the equation to obtain the slopes of the lines of best fit.

The Mean Absolute Deviations and slopes for each of the crime types were then normalized for comparison between the Twitter negative sentiment data and SFPD incident report data. Normalized values were calculated by dividing the values for MAD and Slope by the average negative sentiment score and number of incident reports for each crime type.

	Crime Types	Twitter MAD	SFPD MAD	Twitter Slope	SFPD Slope
0	Person Crime	0.639760	0.874268	0.000227	-0.000002
1	Property Crime	0.796101	0.838314	0.000274	-0.003699
2	Drug Crime	0.720357	0.532240	-0.000116	-0.008776
3	Minor Offense	0.394159	0.749431	0.000570	-0.000803

Table 2: Summary table of MAD and Slopes for each of the four crime types

From the summary table above, property crime has the most similar MAD values for Twitter data and SFPD data. Additionally, the slopes for both Twitter negative sentiment for drug crime and SFPD incident reports for drug crime are negative and are the closest in value.

Analysis of Monthly Trends

To analyze similarities between Twitter negative sentiment and SFPD incident reports over time at a more granular level, the variance and slope of data is calculated on a monthly basis.

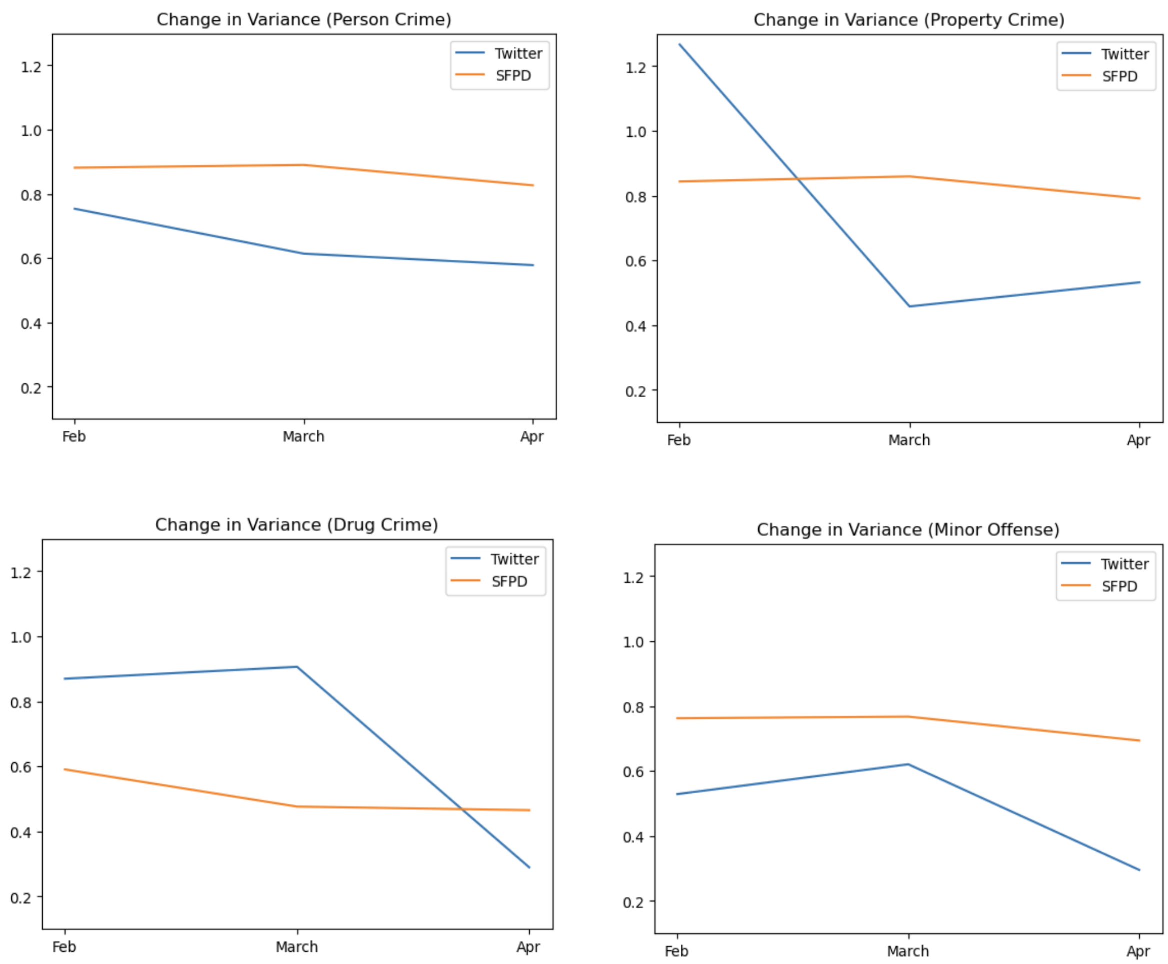

Figure 5: Mean Absolute Deviation (MAD) monthly for Twitter negative sentiment and SFPD incident reports

From Figure 5, changes in variance are similar for Twitter negative sentiment and SFPD incident reports for minor offenses, as fluctuation of both Twitter negative sentiment and incident reports increases for both from February to March, and decreases from March to April. Changes in variance are also similar for person crime, as variance does not shift dramatically across the three months. Variance in property crime and drug crime do not share such similarities, and observe opposite trends across all three months.

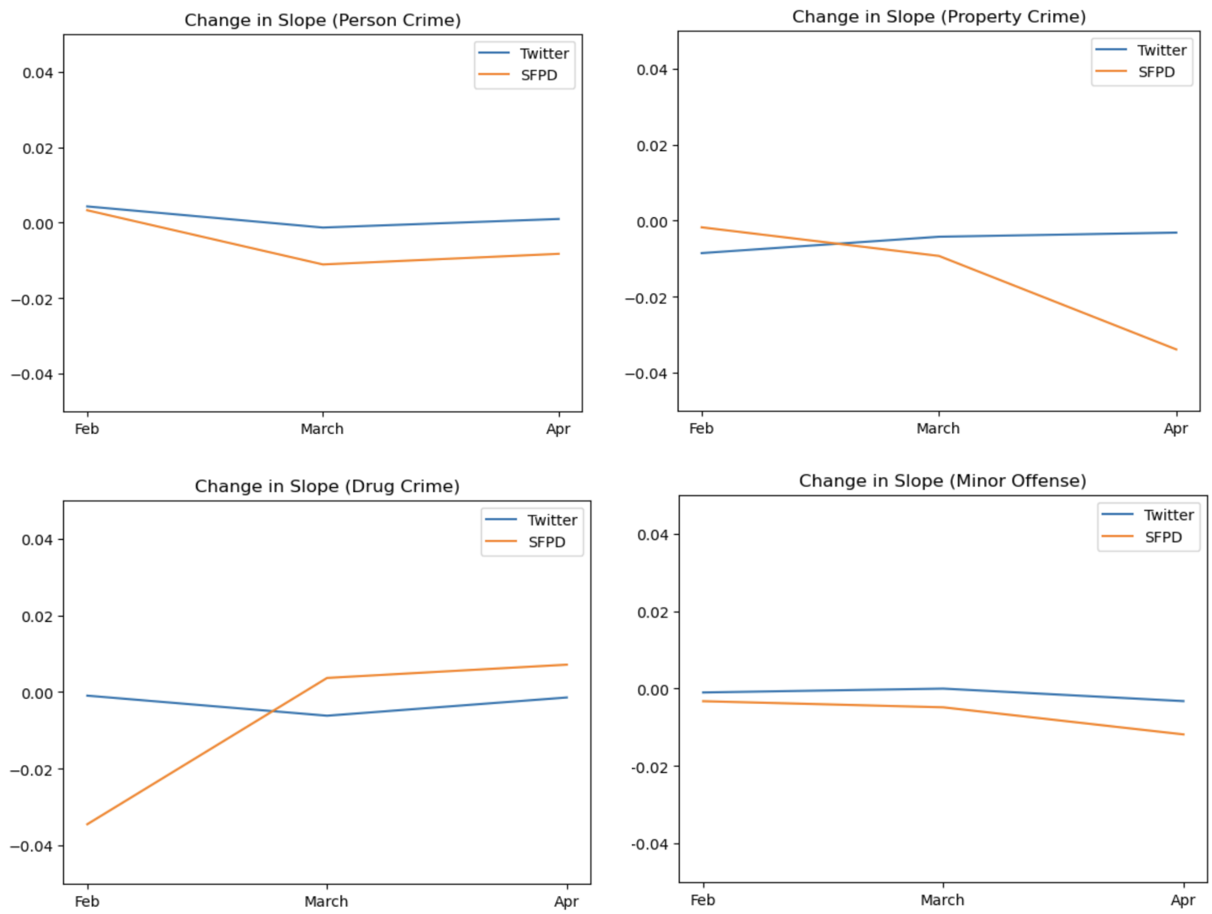

Figure 6: Slope monthly for Twitter negative sentiment and SFPD incident reports

From Figure 6, changes in slope share similarities for Twitter negative sentiment and SFPD incident reports for minor offenses, as negative sentiment and the number of incident reports stays relatively constant through the three months, and decreases slightly from March to April. Additionally, for person crimes, as the number of SFPD incident reports tends to decrease from February to March and increase from March to April, the intensity of negative sentiment also tends to decrease from February to March and increase from March to April. Similar to trends observed in monthly variance, the slopes for property crime and drug crime do not share such similarities, and actually display opposite directional trends.

An analysis of variance (MAD) and slope for the intensity of Twitter negative sentiment and the number of incident reports in SFPD data suggest that users may be more sensitive and responsive to person crime and minor offenses in the city, compared to property crime and drug crime.

Word2Vec Model

To address which kinds of crime (person, property, drug, minor offenses) are more closely associated with fear-related sentiment, vectorization of the Twitter text corpus is performed.

Text Pre-processing

To pre-process the Twitter data before word vectorization, emojis from the Twitter text were removed using regex queries for filtering emoticons, symbols and pictographs, transport and map symbols, and flags. Links were removed from the tweets, using regex query for identifying strings beginning with "http". For hashtags in a tweet, the hashtag preceding the word was removed, and punctuation was also removed from all tweets.

Training the Model

After pre-processing the tweet text, a corpus of tweets is created, a Word2Vec model is fitted to the text corpus, utilizing the Gensim python library for natural language processing. The entire corpus contains 75,119 unique words. We assign the dimensionality of the vectors to be 16, and assign min_count a value of 1, allocating a 16-dimensional vector for each word in the text corpus.

T-SNE

Dimensionality reduction is then performed on the 16-dimensional vectors for all unique words in the Twitter text corpus. From sklearn, we import the TSNE tool to visualize and analyze this high dimensional data, converting these 16-dimensional vectors to 2-dimensional vectors. The gradient calculation algorithm uses approximate optimization using the Barnes-Hut method, which runs in O(NlogN) time, compared to the exact direct-sum algorithm, which runs in O(N^2) time.

To analyze the relationship between crime type and sentiment of fear, 8 fear-related words are identified: fear, nervous, dread, horror, horrified, worry, worried, and shocked. These specific words were chosen as they are close synonyms of fear and are present within the unique words in the Twitter text corpus.

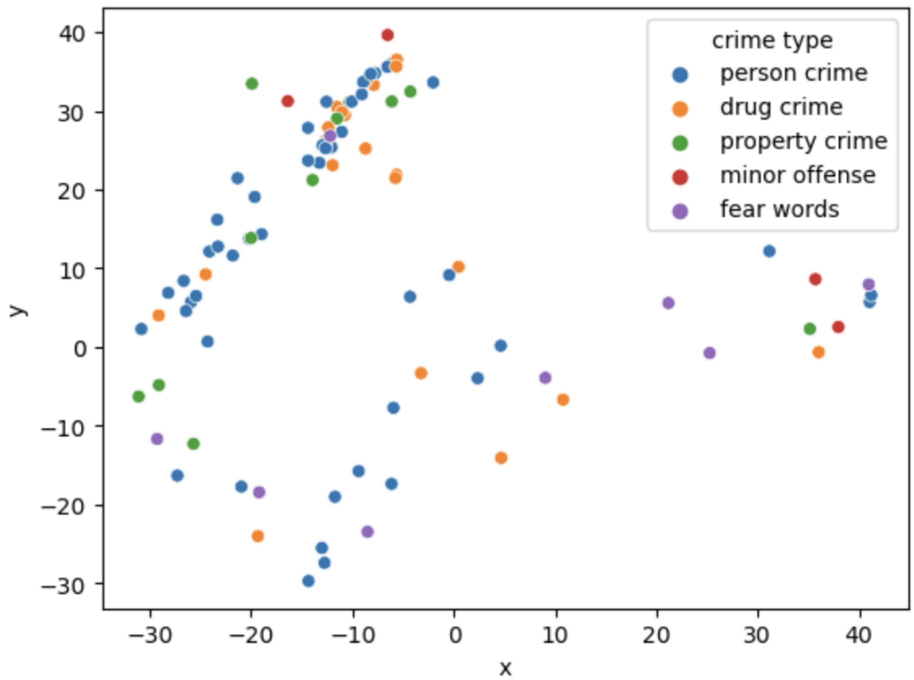

Figure 7: This scatterplot plots the vector positions of crime words identified to extract crime related tweets. The words are color coded based on their associated crime type. The 8 fear words are plotted in purple.

For each of the eight fear words, the five closest crime-related words are identified using euclidean distance. The majority crime type of those five closest crime-related words is then identified:

Fear word	**Majority closest crime type**
Fear	person crime
Nervous	drug crime
Dread	minor offense
Horror	drug crime
Horrified	person crime
Worry	person crime
Worried	drug crime
Shocked	property crime

Table 3: Fear words and corresponding crime types that are most proximate and most common

From Table 3, it is evident that person crime and drug crime are most commonly associated with fear-related words.

Conclusions

From an analysis of trends in negative Twitter sentiment and crime in San Francisco over all three months (February, March, and April), the mean absolute deviation of property crime is most similar across both data sets. There is the largest difference between the mean absolute deviation for negative Twitter sentiment and SFPD incident reports for minor offenses, suggesting that users may not be as sensitive to general trends in minor offense incidents. Additionally, Twitter negative sentiment and SFPD incident reports for drug crimes have the closest negative slope over the three months, suggesting that users may be more sensitive to general trends in drug crime incidents.

An analysis of monthly trends of variance and slope suggests that person crimes and minor offenses follow similar trends, suggesting that users may be more sensitive to person-related crime and minor offenses in their environments. As drug crime demonstrates similar variance and slope to SFPD incident report data over all three months, but does not monthly, this may speak to how user perceptions of drug crimes may not be as sensitive to specific incident reports in the city compared to person crimes or minor offenses, but still may be impacted by general trends. This may be a result of media coverage that does not highlight individual incidents as much as general increases or decreases in drug crime over longer stretches of time. It would take further research to identify the reasons for more broad similarities with drug crime trends. Additionally, more research should be conducted to understand the characteristics of person crimes and minor offenses that may trigger individuals to be more sensitive to those incidents– it is possible that coverage for these kinds of crime may be more heavily focused on in media.

The Word2Vec model suggests that person crime and drug crime is more associated with fear related sentiment. This finding is interesting as it suggests that these crime types may have inherent characteristics that elicit fear in users– perhaps the violent characteristics often associated with person crime are contributing to this association. Additionally, drugs are considered a somewhat taboo topic, and its stereotypical associations may be contributing to its proximity to fear as well. Again, further research is required to understand possible reasons for person crime and drug crime being more closely associated with the sentiment of fear.

Ultimately, this analysis may provide insight for law enforcement to understand how residents of San Francisco may feel about certain kinds of crime, and may be able to better direct their energy and resources into addressing kinds of crime depending on user sentiment. Twitter, being a social media platform, may provide additional indicators of crime– from this analysis, Twitter may be specifically useful for understanding trends in crime for person crimes and minor offenses, at least at a monthly granularity. Additionally, knowing which kinds of crime elicit more fear in individuals may help law enforcement and other citywide efforts and programs to alleviate some of this fear, and to potentially allow residents to live more at peace in their environment.

Limitations

One key limitation of this study is that I only have access to Twitter data from the last 3 months due to changes in the Twitter API. Therefore, I could only conduct an analysis of comparisons between SFPD data within this time frame, and could only analyze user perception within a limited Twitter corpus. With access to more Twitter data, different findings could have been made regarding the types of crime where user fear intensity is most reflective of actual crime, as well as the kinds of crime most associated with fear. Additionally, I utilize the SFPD dataset to represent real crime in the city– these incident reports are not representative of all crime that takes place in the city. In fact, data from the National Crime Victimization Survey conducted by the US Department of Justice indicates that a large number (52.6%) of violent crimes resulting in injury goes unreported to law enforcement (Wu, 2019). Therefore any findings from this analysis using SFPD data regarding user sensitivity to different kinds of crime may not be fully reliable.

Additionally, some decisions made in the methodology offer some limitations. Firstly, the way in which San Francisco related tweets were filtered from tweets extracted from Twitter relied on the presence of a #sf or #sanfrancisco in a Tweet. This filtering method does not account for all San Francisco related tweets that do not include such hashtags, as well as it is possible that some tweets with such hashtags actually do not contain content relevant to the city of San Francisco.

The methods used in extracting crime related tweets also present limitations, as they depend on creating a database of crime-related words and their corresponding crime labels. Both the crime words chosen to include as well as their corresponding crime labels were decided subjectively. As the crime word database does not include all possible crime-related words, but only ones that were considered clearly crime related, all crime-related tweets in the Twitter corpus were not included in the final crime tweet database. Additionally, there were many instances in which a tweet contained a word present in the crime words database, yet the content of the tweet was not in fact crime-related at all. Therefore, findings from this work may not be exclusively reflective of user perception on crime, but may also leave room for the inclusion of non-crime related sentiment.

In learning associations of certain crime-related words with fear, the specific fear-related words chosen for analysis were also chosen arbitrarily. Although the most synonomic words of fear were chosen from the Twitter text corpus for the analysis, it is possible that the inclusion of more fear-related words may alter findings from Word2Vec based fear-association analysis.

Future Work

From a literature review on factors contributing to users' fear of crime, media coverage of different kinds of crime greatly impacts user perception. In the United States, the proportion of citizens who suffer a violent victimization each year is rather small (e.g., U.S. Department of Justice, 1992; Federal Bureau of Investigation, 1993). However, news and other forms of communication about violent crime are far more prevalent, and therefore Americans are very likely to hear about such violent crimes, even if few actually experience it. Therefore, it would be interesting to explore the impact of different forms on media on user perception of various types of crime. This may also provide insight on some of the trends observed through this analysis regarding intensity of fear for specific crime types. It may specifically be interesting to look into media coverage of person crime and drug crime that, from our analysis, seem to be associated more with fear-related sentiment, compared to media coverage of property crime and minor offenses. Additionally, it may be interesting to pursue further analysis of Twitter data, looking into user perception of law enforcement's response to different kinds of crime, to see if perceived inadequacies of law enforcement are reflected in a lack of response or poor response to certain kinds of crime.

References

Arthur, Mac (1999). *Crime Classification Codes.* University of Illinois, Urbana Champaign. http://cda.psych.uiuc.edu/statistical_learning_course/Allcode_manuals/mac_crime_class.pdf

Lal, S., Ranjan, R., Verma, A., & Tiwari, L. (n.d.). *Analysis and classification of Crime Tweets - Researchgate*. ResearchGate. https://www.researchgate.net/publication/340705360_Analysis_and_Classification_of_Crime_Tweets/fulltext/5e99d4c3a6fdcca789206336/Analysis-and-Classification-of-Crime-Tweets.pdf

National Academies of Sciences, Engineering, and Medicine. 1994. *Understanding and Preventing Violence, Volume 4: Consequences and Control*. Washington, DC: The National Academies Press. https://doi.org/10.17226/4422.

Prieto Curiel, R., & Bishop, S. R. (2018, April 17). *Fear of crime: The impact of different distributions of victimisation*. Nature News. https://www.nature.com/articles/s41599-018-0094-8

Prieto Curiel, R., Cresci, S., Muntean, C. I., & Bishop, S. R. (2020, April 2). *Crime and its fear in social media*. Nature News. https://www.nature.com/articles/s41599-020-0430-7

Read "understanding and preventing violence, volume 4: Consequences and control" at nap.edu. Public Perceptions and Reactions to Violent Offending and Victimization | Understanding and Preventing Violence, Volume 4: Consequences and Control |The National Academies Press. (n.d.). https://nap.nationalacademies.org/read/4422/chapter/2

Reeves, J., & | J. R. (2020, February 10). *Survey: People fear cyber crime more than violent crime*. Lawyers Mutual Insurance Company. https://www.lawyersmutualnc.com/blog/survey-people-fear-cyber-crime-more-than-violent-crime

Research and Forecasts. (1980). *Figgie Report on fear of Crime - America Afraid, part One - the general public*. Figgie Report on Fear of Crime - America Afraid, Part One - The General Public | Office of Justice Programs. https://www.ojp.gov/ncjrs/virtual-library/abstracts/figgie-report-fear-crime-america-afraid-part-one-general-public

Ristea, A., Andresen, M. A., & Leitner, M. (2018). *Using tweets to understand changes in the spatial crime distribution for hockey events in Vancouver*. The Canadian geographer. Geographe canadien. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6473699/

Shah, P. (2020, November 6). *My absolute go-to for sentiment analysis - textblob.* Medium. https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524#:~:text=TextBlob%20is%20a%20simple%20library,classifying%20negative%20and%20positive%20words

Wu, D. T., Moore, J. C., Bowen, D. A., Mercer Kollar, L. M., Mays, E. W., Simon, T. R., & Sumner, S. A. (2019, January 1). *Proportion of violent injuries unreported to law enforcement*.

JAMA internal medicine.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6382513/#:~:text=Although%20many%20cities%20rely%20on,resulting%20in%20injury%20goes%20unreported