Analyzing Crime in San Francisco

2020 Vision

Introduction: Crime in San Francisco

San Francisco is a city known for its strong economy and booming tech industry. But in recent years, San Francisco has not been the safest place to live: the overall crime rate in San Francisco is 151% higher than the national average. According to SFChronicle, "homicides increased by 21.4% in San Francisco from March to June of this year," compared to 2019. There is a 1 in 15 chance of becoming a victim of any crime. We want to use this dataset to obtain conclusions about specific factors that correlate to higher levels of crime, which will hopefully inform us of some key insights we can keep during future travels.

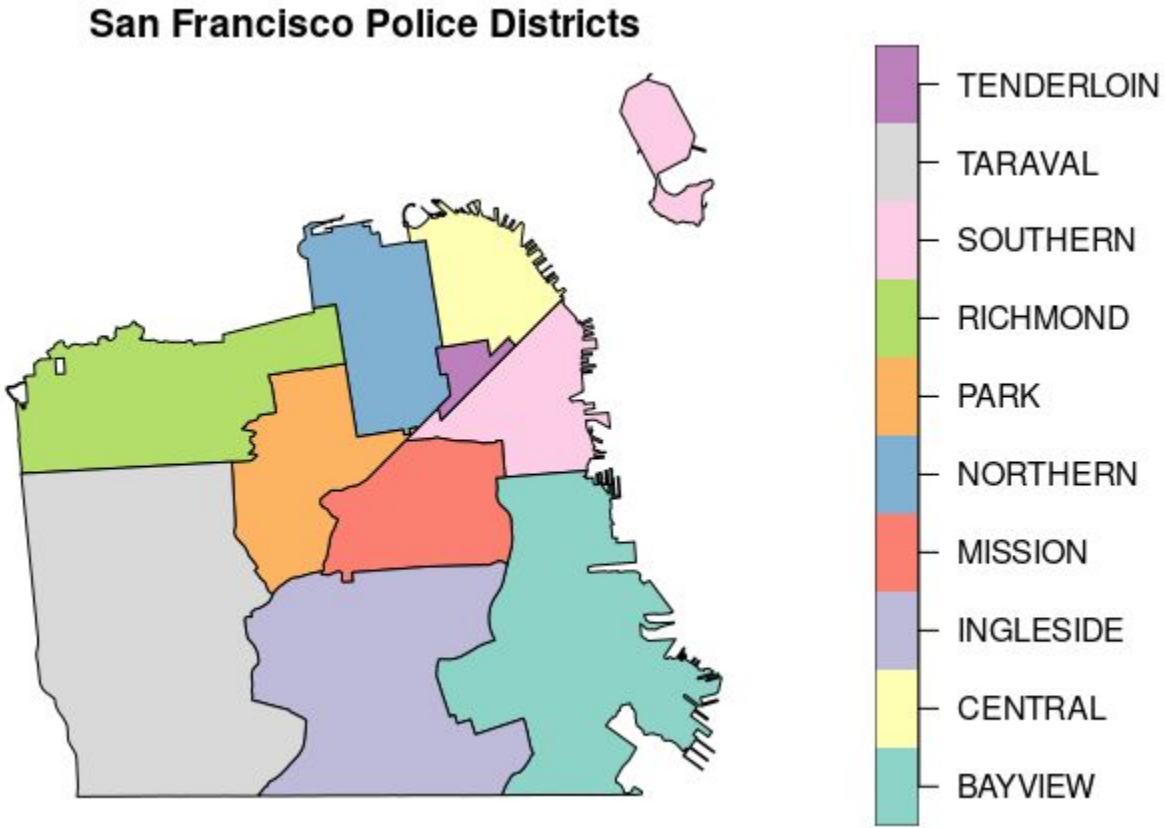
Research Question and Hypotheses

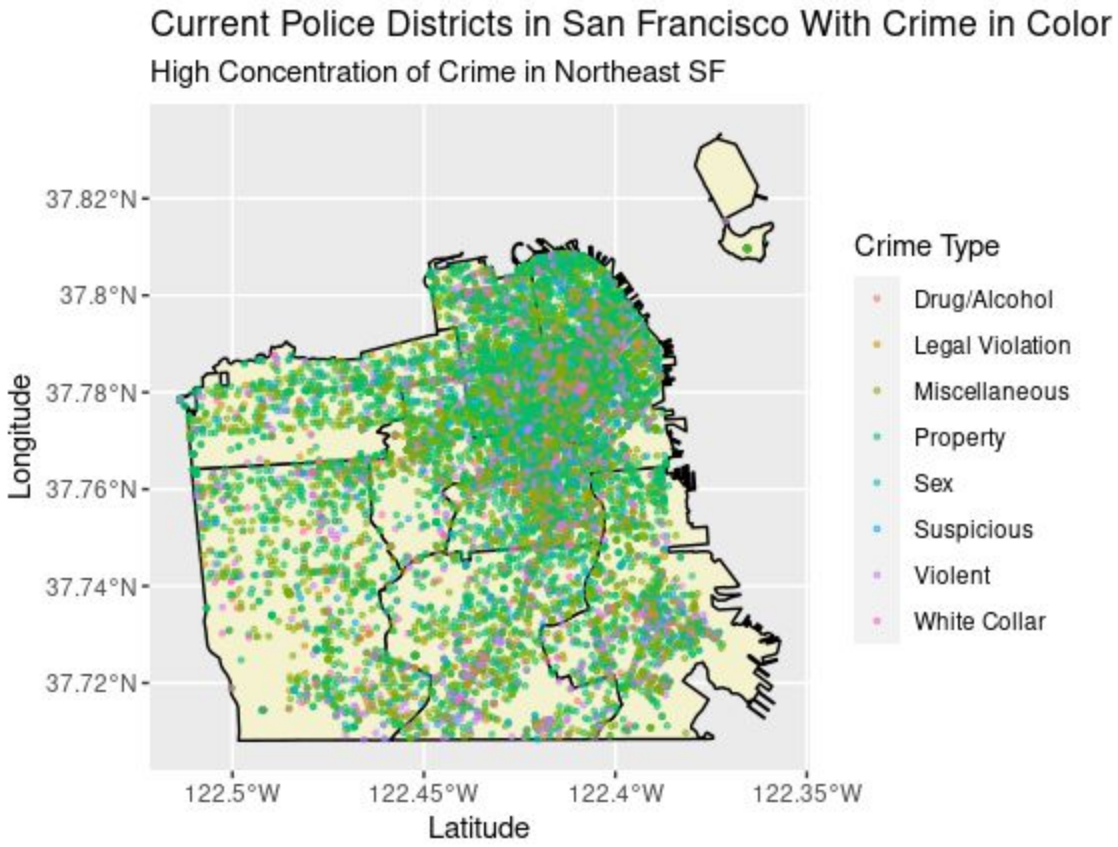
Through our research, we plan to investigate what factors the general population can associate with local crime in order to be the safest while in San Francisco.

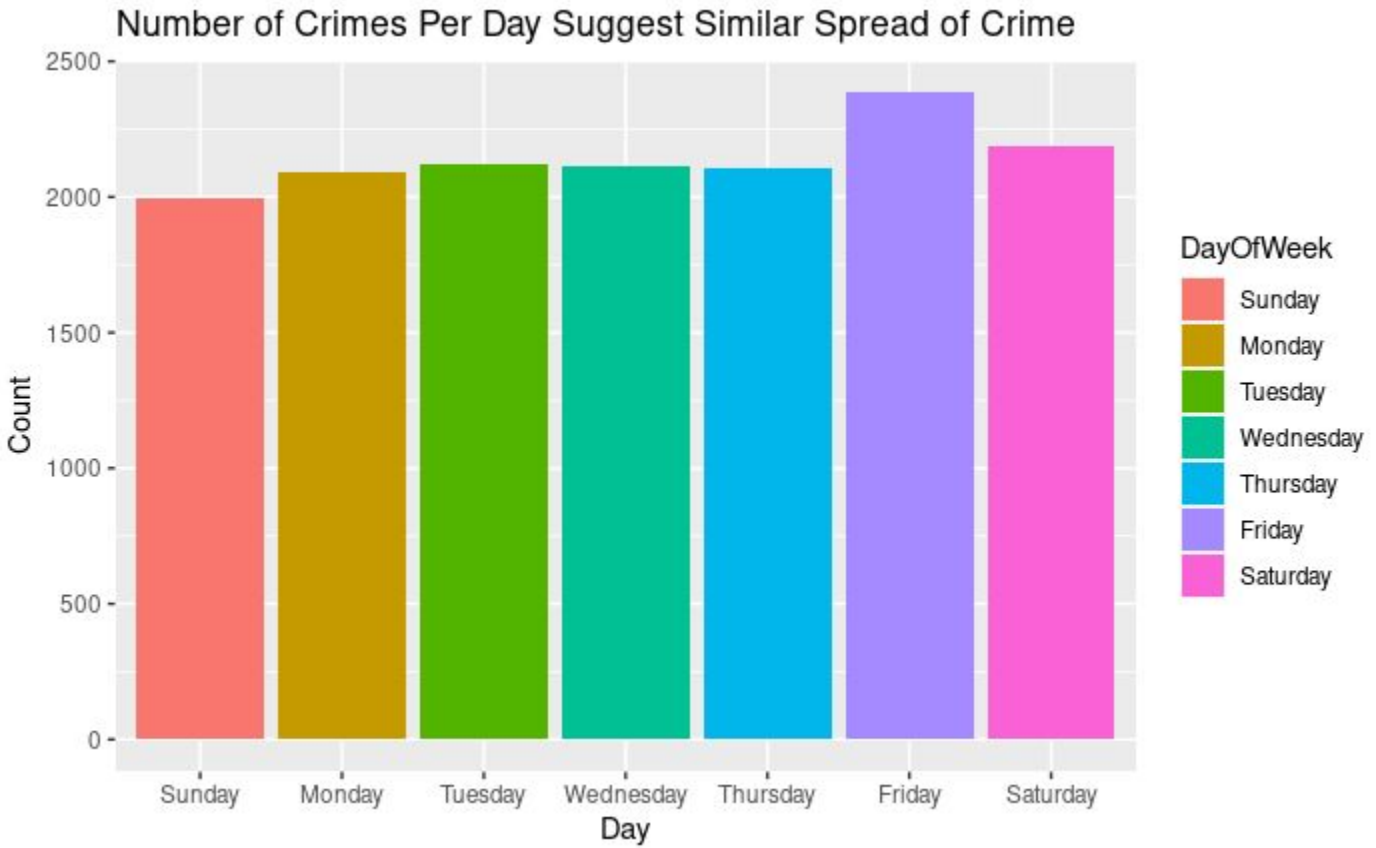
Our main hypotheses are 1) a later time (e.g. nighttime hours) correlates to a higher level or rate of crime and 2) location is correlated to levels of crime. We believe it is important to investigate this question because there likely are policy changes that can be implemented to increase safety throughout the city. For example, if there is a strong correlation between night and rate of crime, then is there a correlation between which night of the week (ex. Sunday night) and rate of crime? With location, are there certain districts that have a specific crime that is common there? By delving further and examining these relationships, we will be able to understand if crime has a specific pattern in SF.

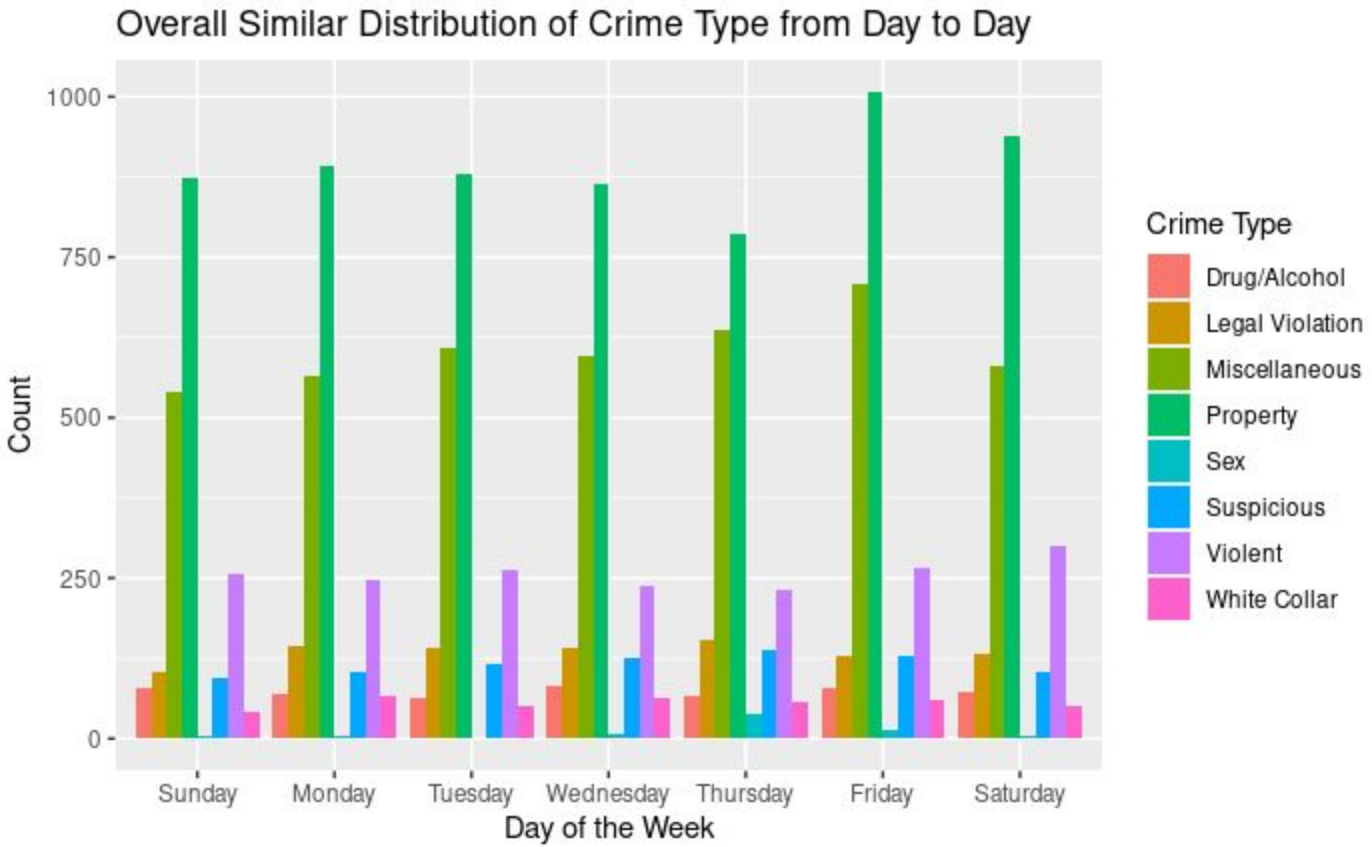
In order to assess these hypotheses, we will look at the following relationships:

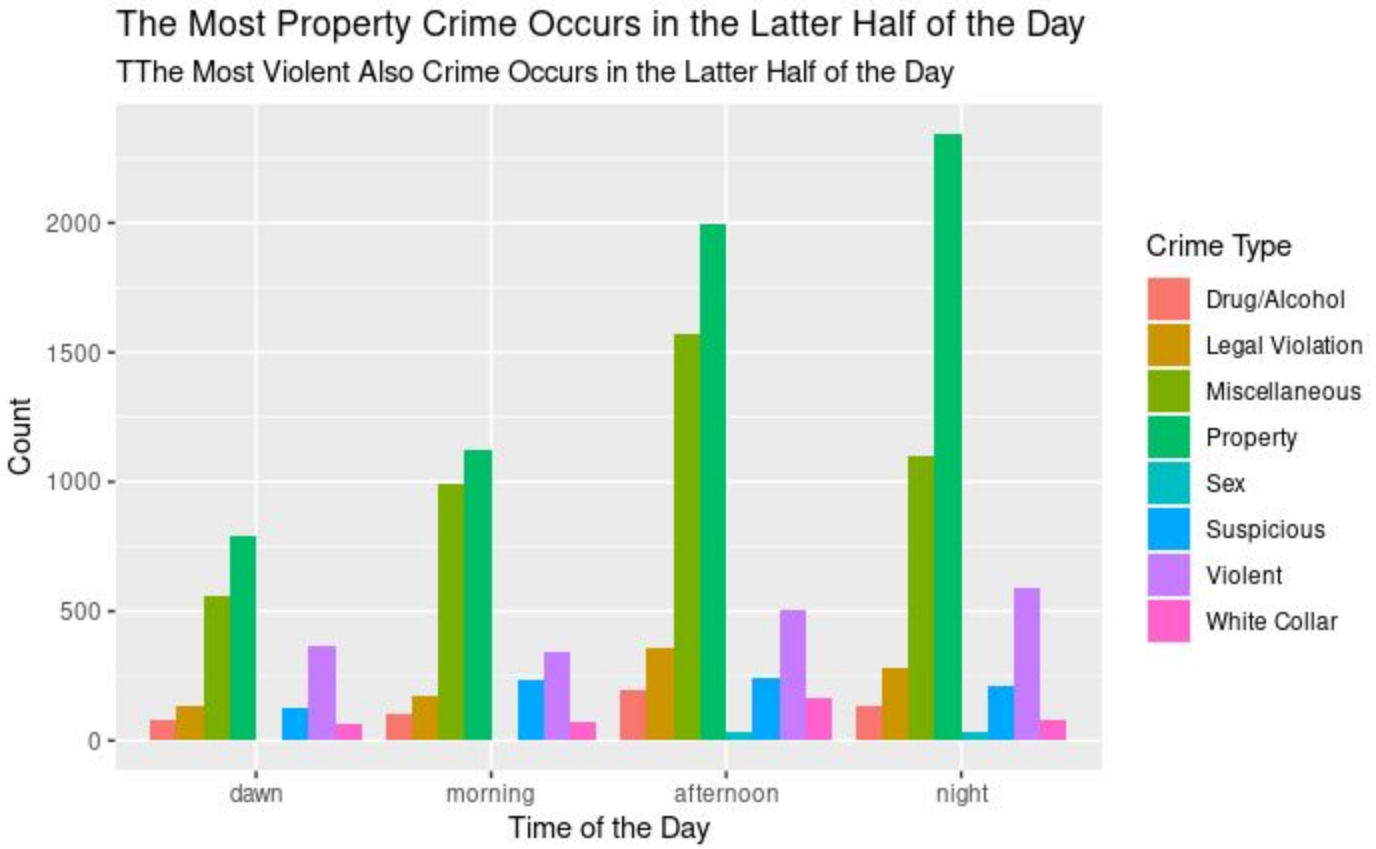
1. Relationship between crime type and time
2. Relationship between crime and time
3. Relationship between violent crimes and police district
4. Relationship between days of the week and crime


San Francisco Police Districts
TENDERLOIN
TARAVAL
SOUTHERN
RICHMOND
PARK
NORTHERN
MISSION
INGLESIDE
CENTRAL
BAYVIEW

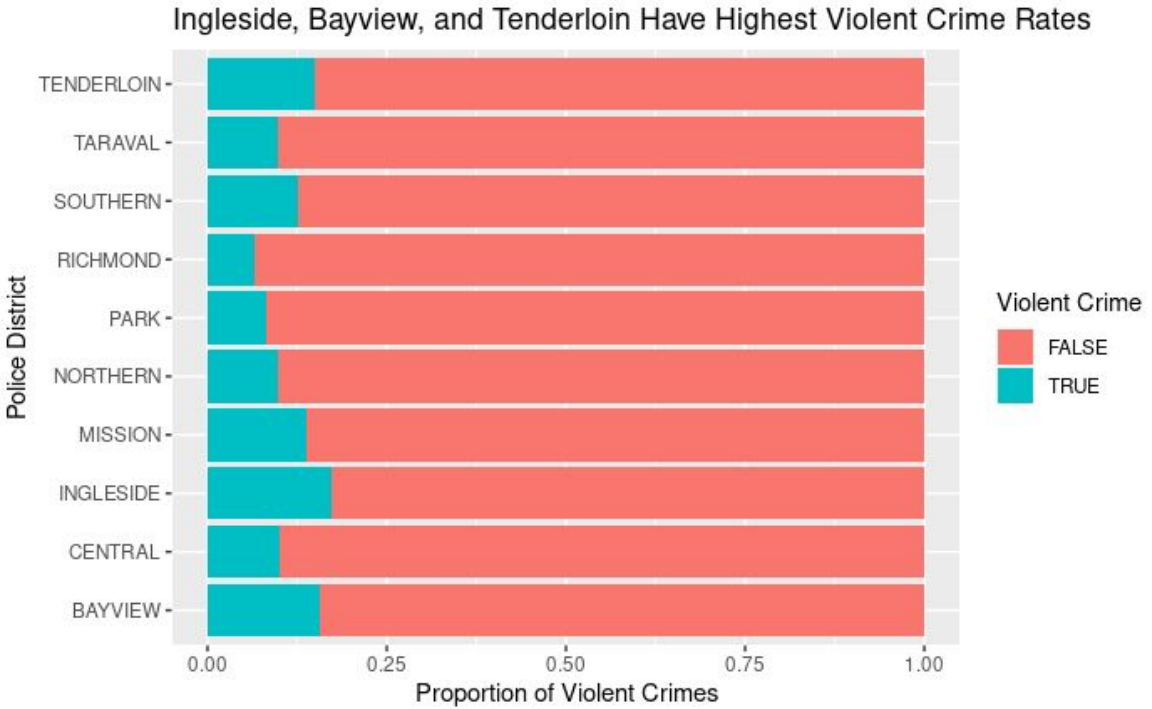

Current Police Districts in San Francisco With Crime in Color
High Concentration of Crime in Northeast SF
37.82°N
37.8°N
37.78°N
37.76°N
37.74°N
37.72°N
Longitude
122.5°W
122.45°W
122.4°W
122.35°W
Latitude
Crime Type
Drug/Alcohol
Legal Violation
Miscellaneous
Property
Sex
Suspicious
Violent
White Collar

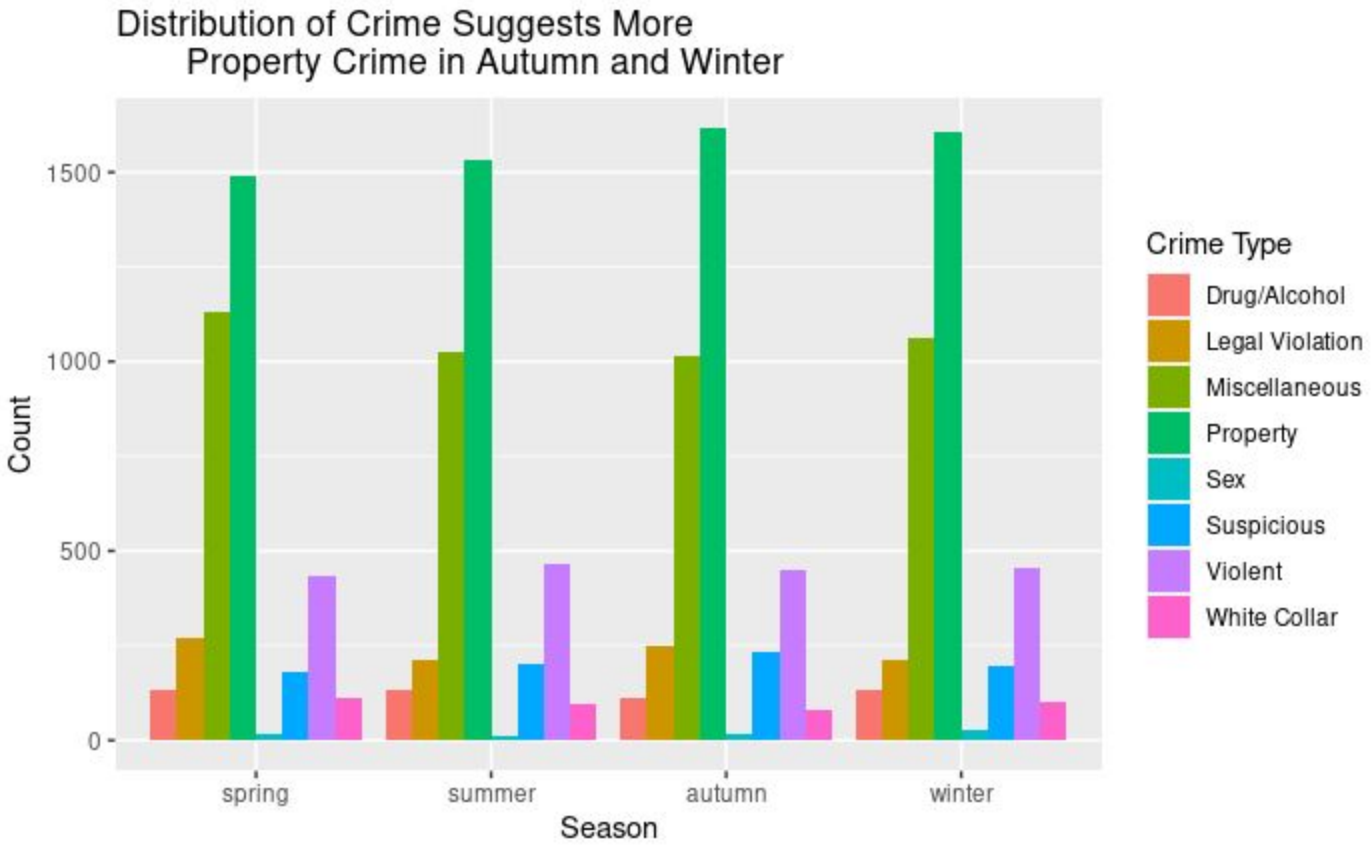

Number of Crimes Per Day Suggest Similar Spread of Crime
2500
2000
1500
1000
500
0
Count
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day
DayOfWeek
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday


Overall Similar Distribution of Crime Type from Day to Day
Count
1000
750
500
250
0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of the Week
Crime Type
Drug/Alcohol
Legal Violation
Miscellaneous
Property
Sex
Suspicious
Violent
White Collar


The Most Property Crime Occurs in the Latter Half of the Day
TThe Most Violent Also Crime Occurs in the Latter Half of the Day
Count
2000
1500
1000
500
0
dawn
morning
afternoon
night
Time of the Day
Crime Type
Drug/Alcohol
Legal Violation
Miscellaneous
Property
Sex
Suspicious
Violent
White Collar

PdDistrict <chr>	n <int>	perc <dbl>
TENDERLOIN	176	17.724068
BAYVIEW	239	16.597222
INGLESIDE	194	15.993405
MISSION	261	13.679245
TARAVAL	128	11.743119
SOUTHERN	291	10.027567
NORTHERN	202	10.004953
CENTRAL	172	9.976798
PARK	77	8.661417
RICHMOND	57	6.934307




Distribution of Crime Suggests More
Property Crime in Autumn and Winter
Count
1500
1000
500
0
spring
summer
autumn
winter
Season
Crime Type
Drug/Alcohol
Legal Violation
Miscellaneous
Property
Sex
Suspicious
Violent
White Collar

Discussion: What Matters for Safety

Not all factors play a substantial role in crime:

- Day of the week
    - Each day had a similar spread with property crime being the highest followed by miscellaneous crime then violent crime

HOWEVER, we did find a relationship between time of day and type of crime committed

- As a result, we created a logistic model to calculate the log-odds of whether a violent crime occurred with the predictors of day of the week and time of day.
- We created a similar model for property crime as well through t-tests

Most cities will have likely have crime peak during the night and during the weekends because more people will not be home.

Discussion: Breakdown of Violent Crime

- The police districts of Tenderloin, Bayview, and Ingleside have the highest percentages of violent crime
- Bayview, Northern, and Southern that have the highest total number of violent crimes
- Park and Richmond were both consistent in having the lowest numbers of total violent crimes as well as proportion of violent crimes
- Noting the success of these districts in maintaining low levels of crimes, it could be beneficial to restructure other districts to mirror their practices

Conclusion: Where to Go From Here

From a policy standpoint:

- Government leaders in San Francisco should consider having additional police on duty during the times when crime is more eminent (afternoon and night)
- Hire more police trained in Larceny, Theft, and Assault, as they were the most prominent from the graphs
- Police districts with higher violent crime rates should have more armed officers, while areas more heavily burdened with non-violent offenses could be patrolled by unarmed officers

Again, we understand that we cannot extrapolate our analysis to every city; however, our conclusions will be generalizable to similar cities to a moderate degree (other cities with similar infrastructure and economic conditions are more likely to utilize the analysis we've found)

Conclusion: Where to Go From Here

If we were to continue work on the project, we would add to our analysis by:

- Introducing data from different cities that are comparable to San Francisco. It would be interesting to see the parallels in crime rates, as expanding the population of interest to citizens in multiple cities would give a better picture of how cases of crime occur differently by region, state, country, or population density (urban vs. rural)
- Adjusting for additional potential confounding variables to improve the accuracy of our analysis and models
- Assessing population numbers for each police district (having a larger population size would likely contribute to greater numbers of crime, even if per capita crime is lower. This information is not present in the dataset we used, but would be necessary to extrapolate a greater conclusion regarding which police district is most dangerous. Given that factors such as poverty level and unemployment rates are main drivers for crime [1], it would be valuable to Assess population numbers for each police district).
- Finding the differences in these factors for districts with more and less crime so that next steps can be taken to lower crime rates. For example, should a future study conclude that Park's public education system has higher test scores than that of Bayview, improving schools could be the best step for mitigating crime.