**San Francisco**

**Property Value Vs. Crime Rates**

Group 3

Ezgi Booth,

Madison Hanson,

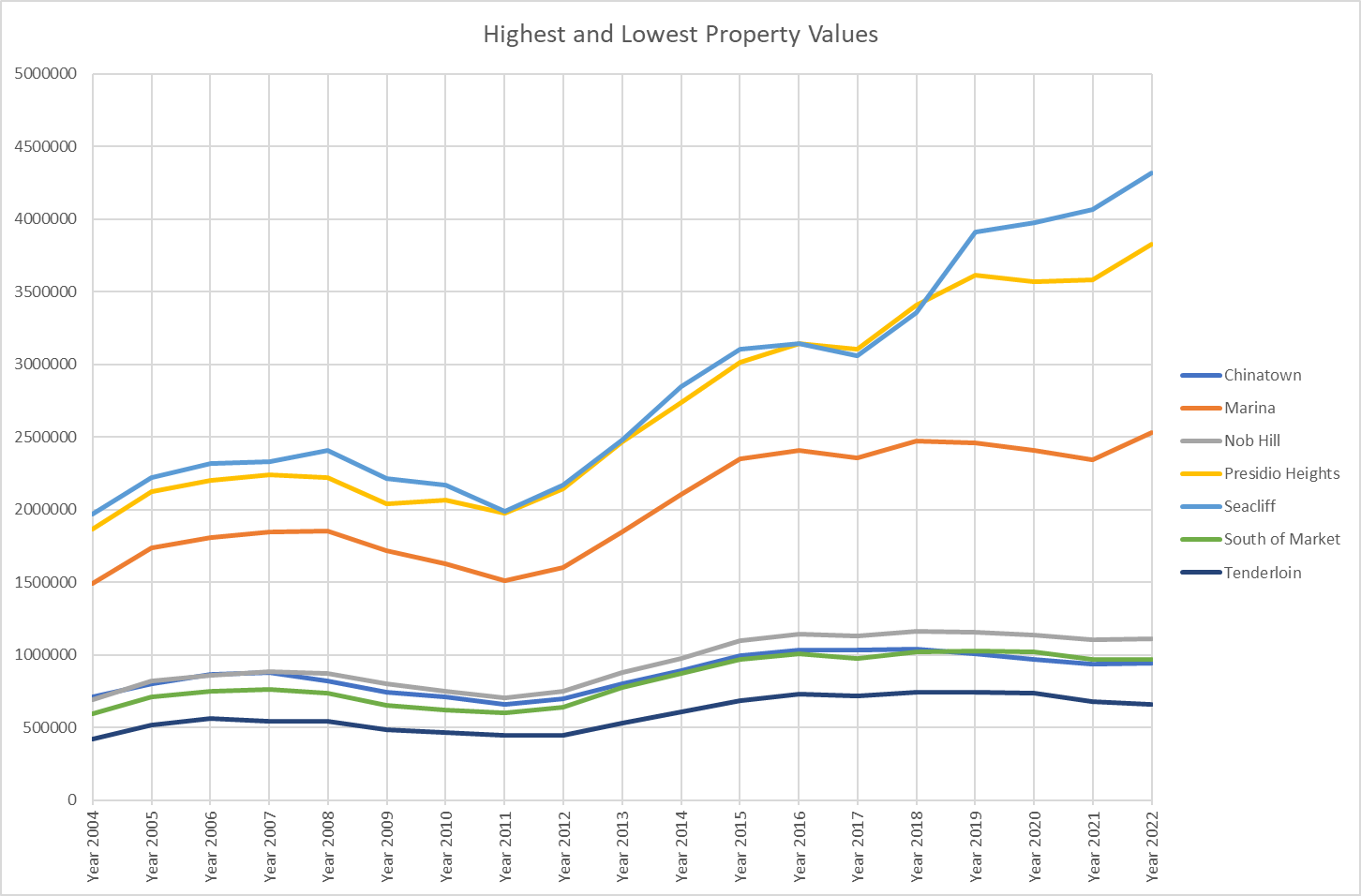
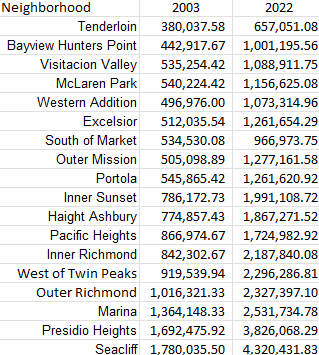
Matthew Ng,

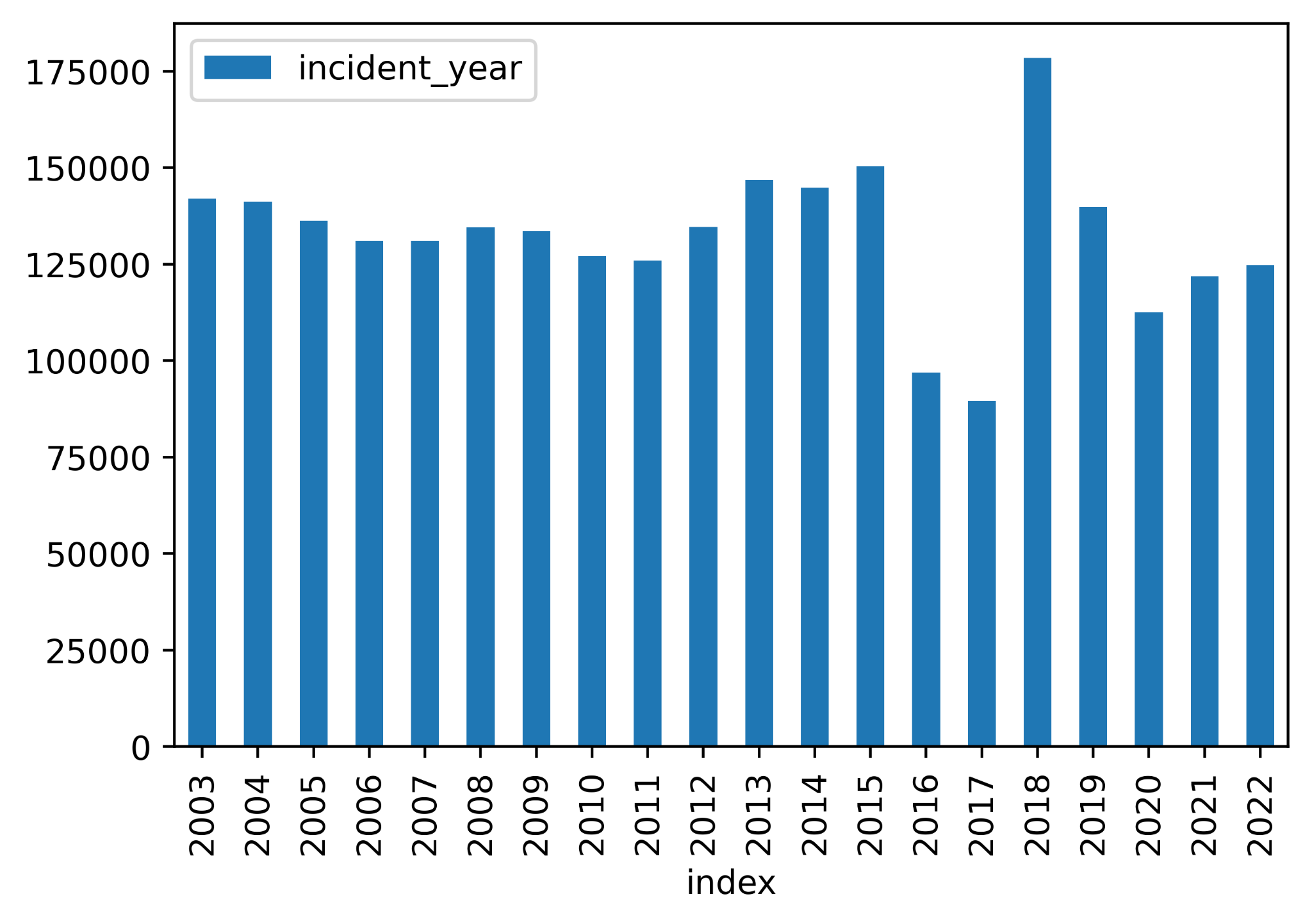
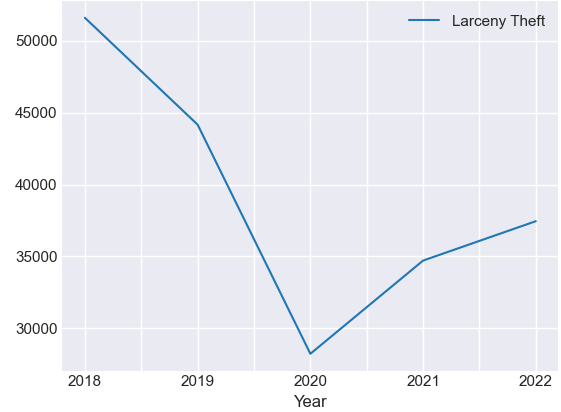
Michael Banning,

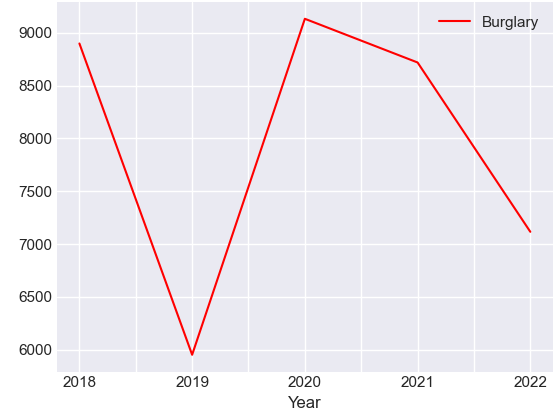
Rashed Alkanawi

How much does a neighborhood's crime affect its desirability to live in? Though there are many contributing factors to housing prices, we shifted focus on the impacts that arise from crime rates. A [study](https://core.ac.uk/download/pdf/217505054.pdf) on how the proximity of crime impacts housing prices raised an interesting question concerning socioeconomic implications–does poverty cause crime, or does crime create poverty around it(Thompson, 2018)? In this project, we are assuming that property value is negatively impacted by crime rather than crime being a product of surrounding property values, meaning San Francisco property values are negatively correlated with higher incidents of crime.

We gathered housing data from Zillow’s Home Value Index and crime data from San Francisco’s police department via DataSF’s API. After cleaning and analyzing each dataset separately, we merged the two and conducted statistical analyses of San Francisco’s housing prices and crime data for the past 20 years (from 2003 to 2022) to determine correlation and statistical significance. Both sets provided information on the neighborhood level.

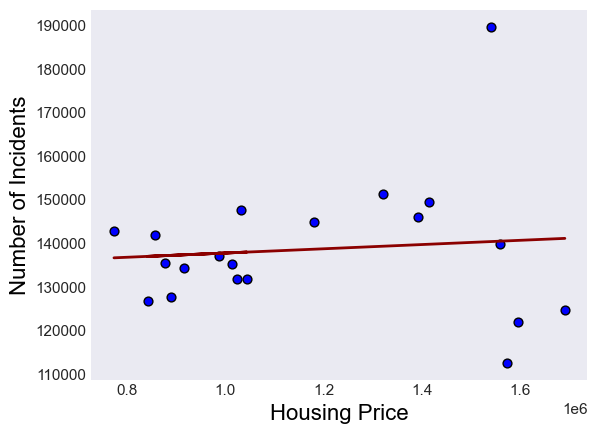
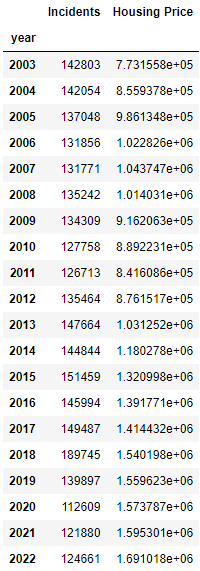
Like any major city there are clear housing price divides between neighborhoods. San Francisco is no different, with regions remaining consistent in the higher or lower end of this observation, continuing to broaden the gap. To understand how the housing rates have changed over the last 20 years we utilized the average housing cost per zone per year. Regions such as Seacliff, Presidio Heights, and Marina have stayed on the upper end of this scale while Civic Center, South of Market, Tenderloin, and Nob Hill have remained at the bottom. Regardless of the property value ranking of the region, each saw at least a double increase from 2003 to 2022. It is clear that the city’s ability to peek prospective homeowners interest is strong, so now we need to see the patterns within its crime rates.

Populating the San Francisco crime incident reports through API provided clear data such as the exact latitude and longitude of the reported incident and category type.The most frequent types of crime in all regions of San Francisco involved some form of theft or burglary. Over the last 20 years this has held true with larceny making up 23.7% of crimes committed with the next distinguishable incident report being assault at 7.48%. Over this time period the mean number of crimes reported was 132,194.75. Covid-19 seemed to act as a preventative force as there was a 20.63% decline in the number of reports in 2020. As the city transitioned into its post pandemic era, a 7.01% increase in crime occurred in 2021.

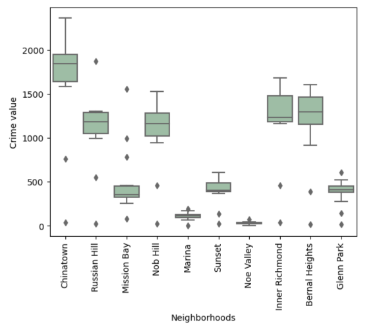
 Larceny theft saw a -36.12% change from 2019(44,173) to 2020(28,217) due to the shift of in-person work becoming remote or terminated as well as the lack of tourism. The following year showed an increase of 23%(34,708), proving the suppressing effect of the at-home order on larceny. The opposite can be said when focusing on burglary. An increase of 53.49% from 2019(5,951) to 2020(9,134) occurred, with a decrease of 4.53% in 2021(8,720). These burglaries affected mainly vacant commercial and office buildings. Over the studied 20 year span of time, the most crime-ridden regions in San Francisco were Civic Center, South of Market, Tenderloin, Nob Hill, and Mission District.

Now that we have analyzed the data for the housing market as well as the crime rate in San Francisco for the past 20 years, we can begin to look at whether or not there are relationships between the two, and to do that, we conducted various statistical analyses.

For the first correlation analysis, while we still looked at housing prices and the number of reported incidents from 2003 to 2022, it is important to note that we used the sum of all incidents across all categories, and for the housing prices we used the average price across all regions, within a given year. Plotting the data, we saw that as housing prices increase, the number of reported incidents remain fairly clustered in the lower price region; however, as we approach properties with higher values. In this case the $1.4 to $1.7 million range, we saw that the data points are less clustered, with some years exhibiting a relatively low number of incidents and even one exhibiting a lot higher than usual activity, which is an outlier. Nonetheless, when performing a linear regression, we saw that the r-value–the correlation coefficient–is 0.01, which actually indicates a very weak correlation between the average housing price and the number of incidents reported in San Francisco.



For our second correlation analysis, we chose to look at a specific year (in this case 2022 as it is the most current) and conduct a more in-depth analysis that accounts for several regions of San Francisco. For the regions, we selected the five highest and five lowest valued neighborhoods and compared them against their crime values. We chose to go with ANOVA as it allows for a comparison of more than two groups at the same time to determine whether a relationship exists between them. The ANOVA test resulted in an f-value of 43.8 and a p-value of 1.25e-37, which is well below 0.05, meaning the data is statistically significant. Running a correlation analysis on the same data reveals an r-value of -0.57, indicating a moderate negative correlation–meaning home values tend to be higher where crime values are lower, and vice versa.



While there is weak to moderate correlation when comparing housing prices to the number of reported incidents, both models have p-values well below 0.05, which indicates that the difference between the means is unlikely to have occurred by chance.

The obstacles that arose during the analytical process are important to consider when looking to make conclusions. The data sets available consisted of differing region definitions. Zillow recognized 108 unique neighborhoods, while DataSF saw 39. In order to combat this issue we found the exact longitude and latitude of each region in the more diverse data set and determined which region they would be considered a part of within the 39 region data set. This allowed for a consistent baseline for region analysis.

The general San Francisco area has seen no real signs of slowing down in terms of its average home prices and crime count over the past 20 years. The two seem to not be very dependent on one another on a larger scale. Although, when focusing on a specific year and the maximum and minimums of the property value spectrum, a moderately negative correlation with incident reports can be drawn. This confirms the original assumption that high crime rates are likely to produce lower home prices within the area. There is a price to be paid to feel safe in your home.

References

Thompson, Samuel D.. "How the Proximity of Crime Impacts Housing Prices: A Hedonic Pricing Study of Inner-Loop Houston, TX.." (2018)https://digitalrepository.unm.edu/geog\_etds/41