Exploratory Data Analysis

Data Analytics Results

Figure 1 shows the comparison between the house price values and the median household income values in each county. Looking at the figure, we can see the pattern where the house price mainly varies according to the median household income. When the median household income increases, the house price in the same region is also more. For example, income and house prices are higher in Alameda county when compared to Butte county, which has relatively low income and house prices.

Figure 1

Bar and Line Graph Comparing House Price and Median household Income

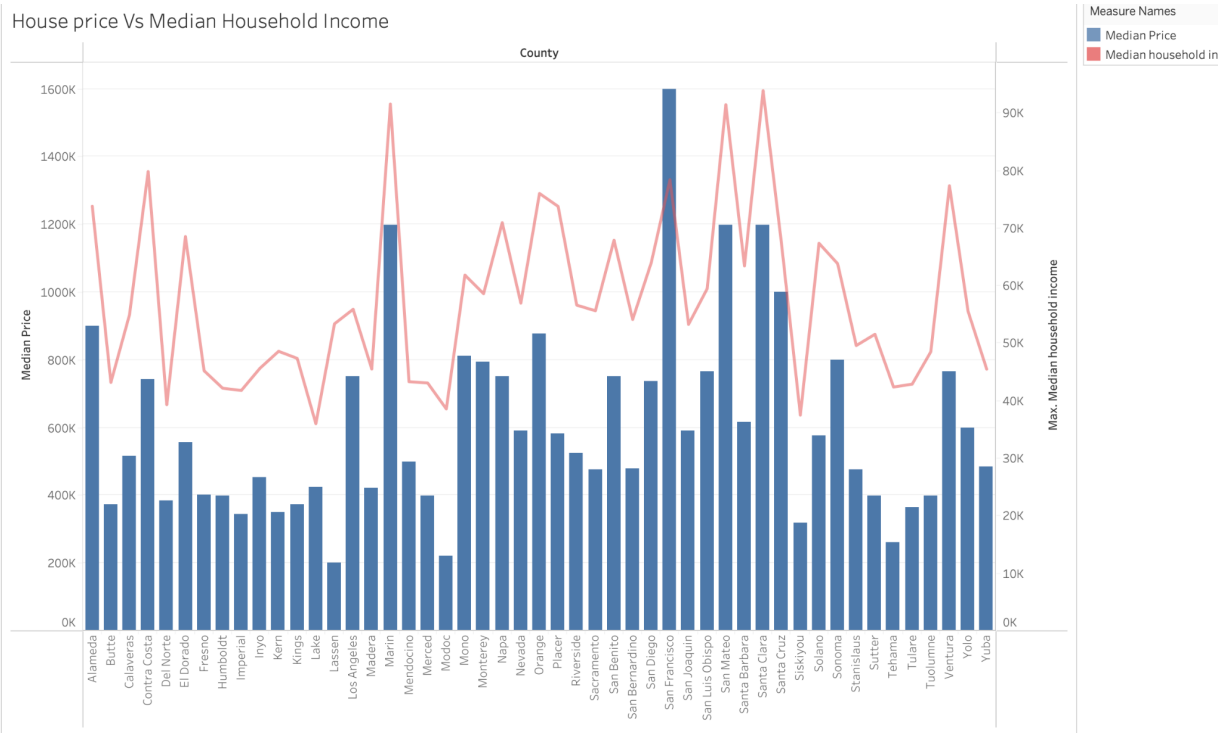
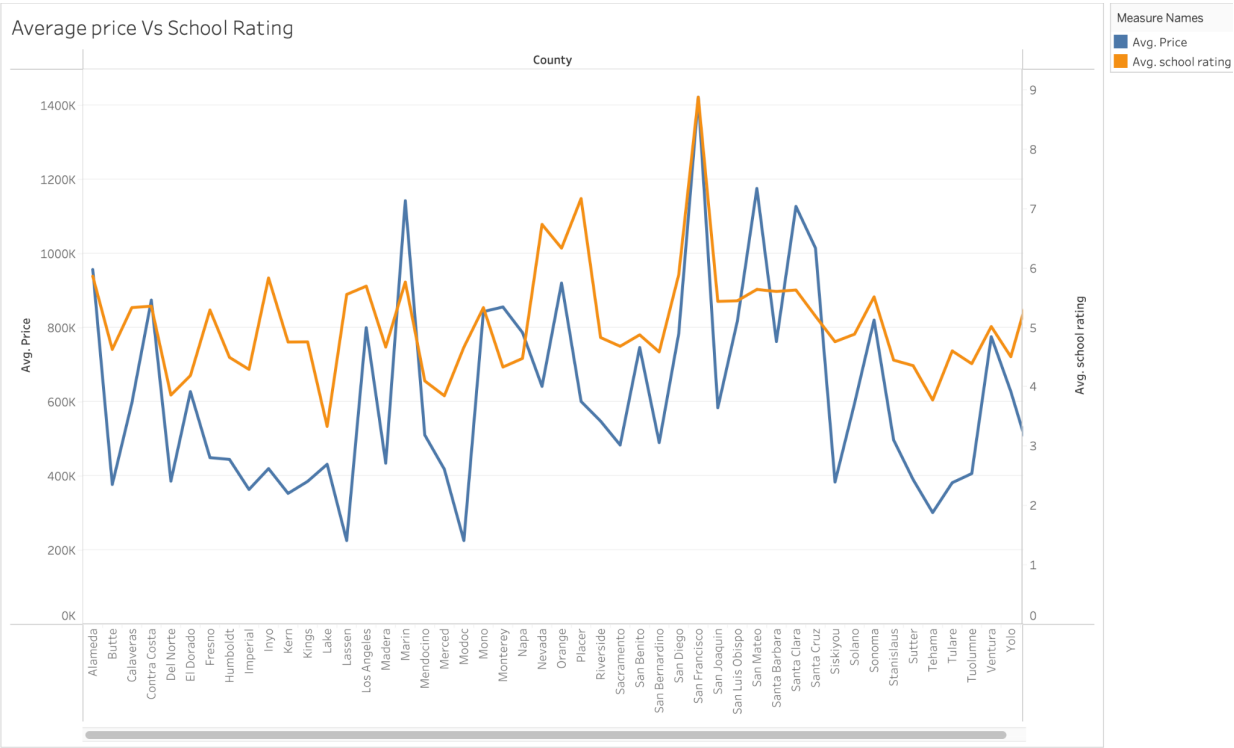


Figure 2 shows the comparison between the average price and the school rating in each county. By observing the trend in the figure, both the line graphs follow a similar trend. When there's an increase in the average school rating, the price also increases. This graph shows that the school rating is also one of the deciding factors in determining the house price in a particular area. For example, San Francisco county has the highest school rating, and houses are expensive compared to other counties.

Figure 2

Line Graph comparing Average Price and School Rating



To understand the relationship between the descriptive features with the target feature "price," a dashboard was created in Figure 3. The left figure displays the average overall school ratings(elementary, middle, high) with a color channel and the 'price' depicted by size. These big data visualizations of the entire state may appear a little clumsy because they are a lot of data.

Still, when the same figure is zoomed in, as represented in Figure 4, we see a more detailed perspective of the cities. We can now notice an interesting difference between Palo Alto and east Palo alto. Even though they are adjacent, they have drastically different school ratings. Irrespective of that, we can observe that the house prices remain almost the same in both regions.

Figure 3

Dashboard Representing Relationship of Schools and Crimes With House Prices

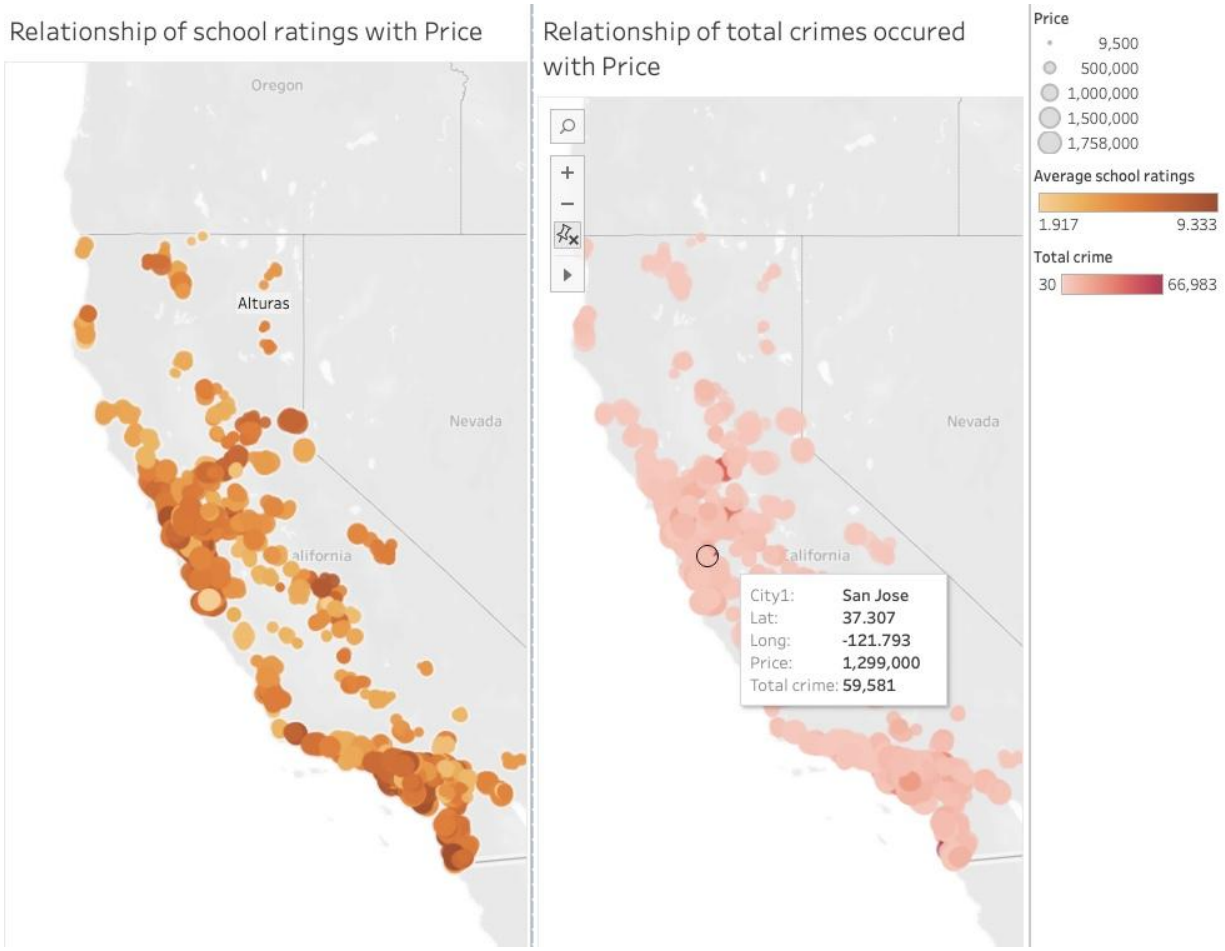
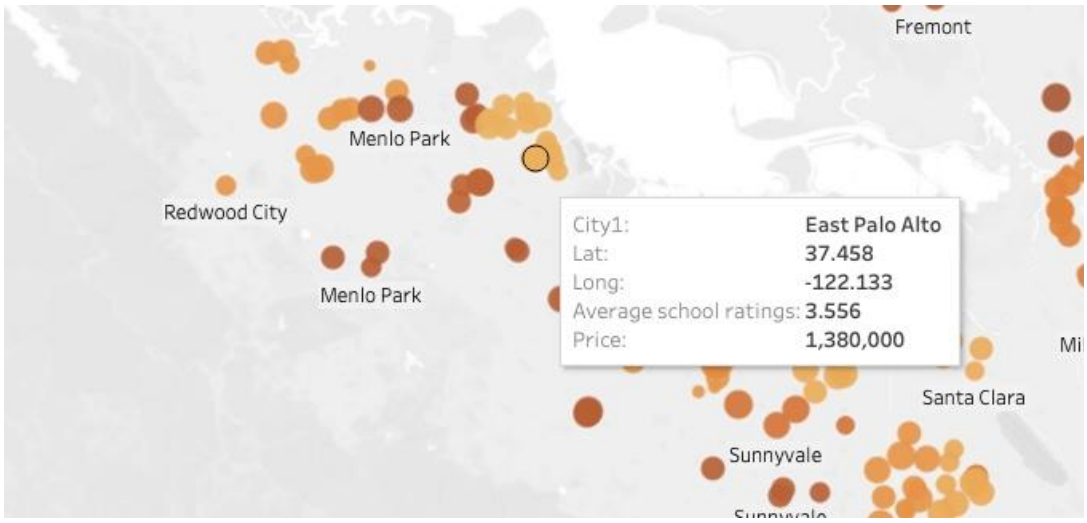


Figure 4

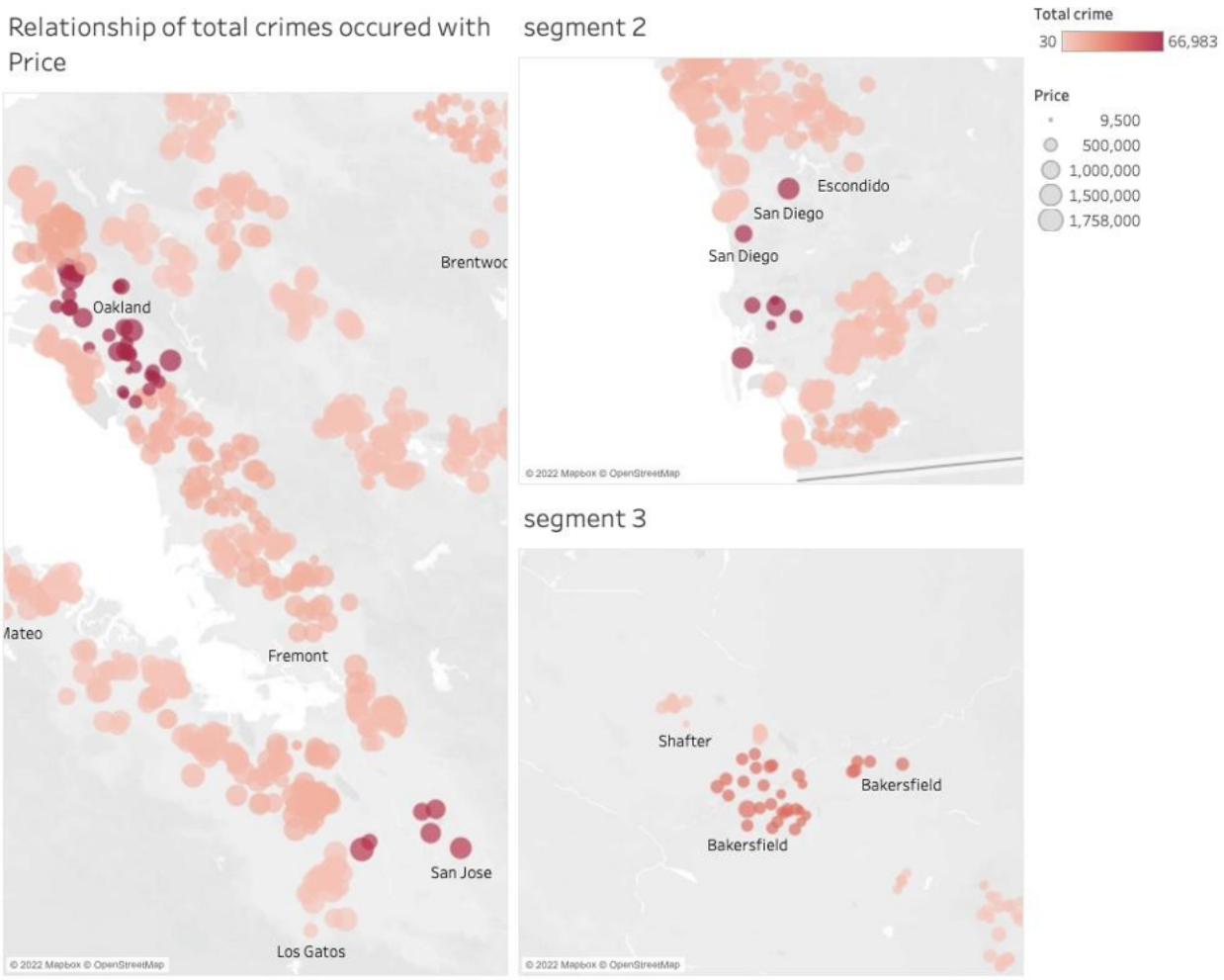
Detailed View of School Ratings With Price



Similarly, the right geospatial map in Figure 5 represents the total crimes in a city along with the average prices. To find unbiased results, we removed the outliers dominating the total crimes feature. The sum of crimes is represented using color, and the price is represented using size. The dark color and big size points represent higher crime rates with higher-priced properties. This kind of cluster can be observed in Oakland, San Jose, San Diego, and Bakersfield cities, which are the highest crime clusters when mapped.

Figure 5

Detailed View of the Crime vs. Price Map



The following Figure 6 illustrates the counties with the most expensive house prices in California. We have used a diverging color scheme, which tells us that the counties near dark blue have costly houses (Sonoma, San Mateo, Santa Clara, and Monterey). Also, the counties near the white region have neutral prices, and the counties near the red region have relatively low house prices compared to the other two.

Figure 6

The Counties in California and Their Housing Price Rates