

California Housing Prices Visualization

By Pengfei Liu

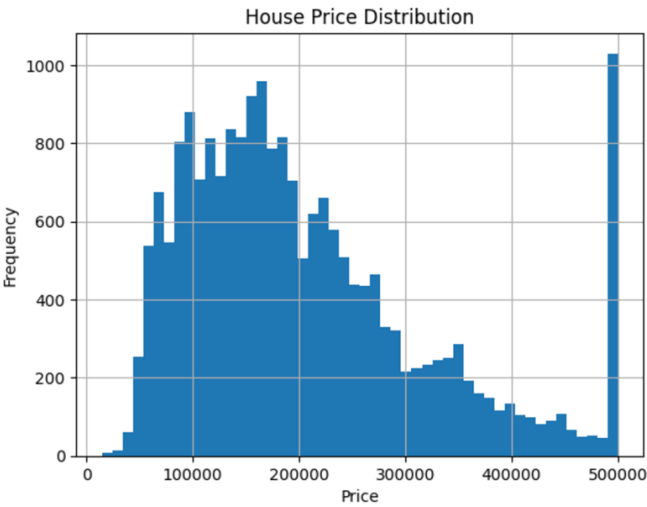


Figure 1: House Price Distribution

Figure 1 shows a notable spike at the upper limit (\$500,000), which suggests that the data might be capped or truncated at this value. The majority of house prices fall within the range of \$100,000 to \$300,000. The distribution is right-skewed, with a long tail toward higher prices. This indicates that while most houses are in the lower-to-middle price range, a smaller number of houses are priced higher.

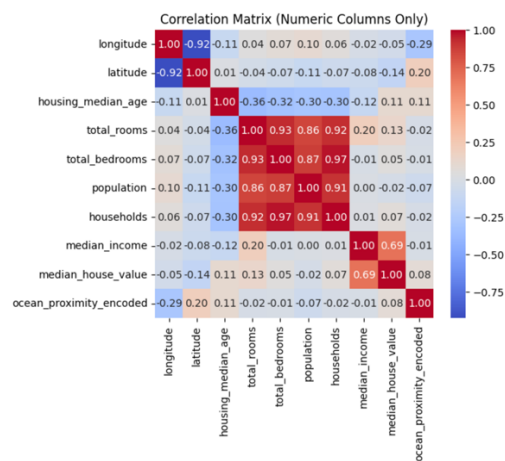


Figure 2: Correlation Matrix

Figure 2 is a correlation matrix heatmap, which visualizes the relationships between numerical features in the dataset. Strong correlation indicates that areas with more total rooms also tend to have more bedrooms, which is logical. The plot also Indicates that higher income levels are associated with higher house prices, a key feature for predicting housing prices. latitude and longitude (-0.92) reflects a geographic trend since these values represent the physical location of houses in California.

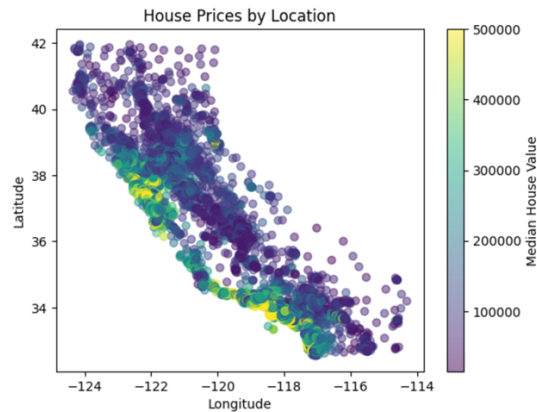


Figure 3: House Prices by Location

Figure 3 visualizes house prices by geographic location in California using latitude and longitude. The yellow and green dots (indicating higher house prices) are clustered along the western part of the map (closer to the coast), particularly near central and southern coastal areas (e.g., Los Angeles, San Francisco). This aligns with the fact that coastal areas in California tend to have higher property values. Purple dots dominate the eastern and inland areas, indicating lower house prices further away from the coast.

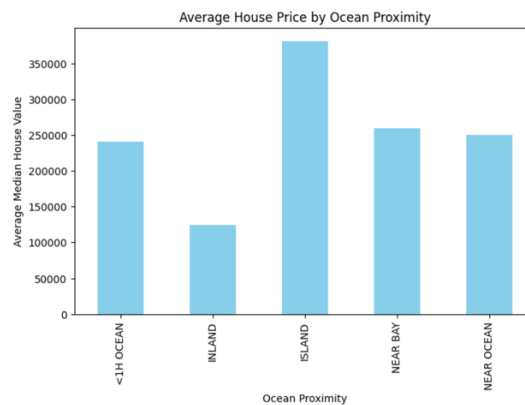


Figure 4: Average House Price By Ocean Proximity

Figure 4 shows ISLAND has the highest average house prices, exceeding \$350,000. This may be due to the scarcity of housing and the exclusivity of island locations. <1H OCEAN and NEAR OCEAN also have relatively high house prices, likely because of their desirable locations near the ocean. Houses closer to the ocean, bays, or on islands generally command higher prices due to demand, exclusivity, and amenities.

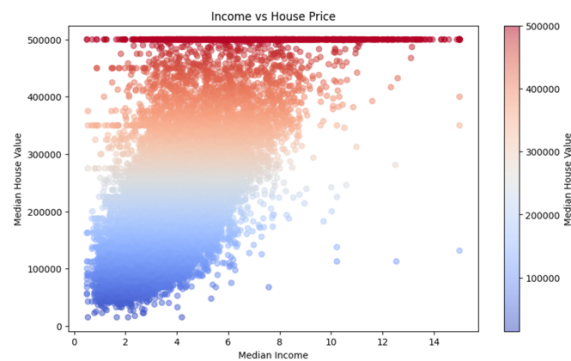


Figure 5: Correlation Between Income and House Price

Figure 5 visualizes the relationship between median income (x-axis) and median house value (y-axis). As median income increases, median house values generally increase, indicating a strong positive relationship. The plot uses color encoding to represent the intensity of the house value, with higher values shown in warmer colors (red/orange) and lower values in cooler colors (blue). As median income increases, median house values generally increase, indicating a strong positive relationship.

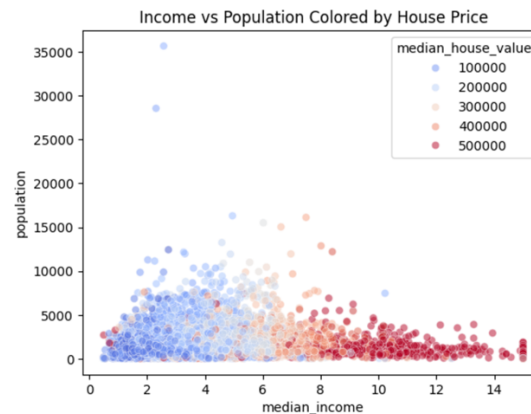


Figure 6: Income VS Population Colored by House Price

Figure 6 visualizes the relationship between median income (x-axis) and population (y-axis), with color indicating the median house value. Per the plot, higher incomes (above 6) are strongly associated with higher house prices, as indicated by the dominance of red-colored points in this region. Most of the data points cluster below a population of 10,000, with very few regions exceeding 20,000 in population. There is no clear relationship between population size and house value; areas with varying populations have both high and low house values.

Importance: Analyzing California house prices is an important topic because housing is a fundamental human need, and California represents one of the most dynamic and complex housing markets in the United States. The provided plots collectively reveal key insights into how socioeconomic factors, geographic features, and population dynamics drive housing affordability and accessibility. Based on all the plots, the California housing market is deeply influenced by income levels, geographic location, and proximity to the coast. High-income areas and coastal regions dominate the market with higher house prices, while inland areas remain more affordable but less desirable. These analyses will provide reliable information for people who are planning to buy a home in California.

Data and Method: The dataset I used for this assignment came from open source Kaggle California Housing Prices. All of the graphs are generated by Python with different plotting packages like seaborn, pyplot, and matplotlib.

Github Link: <https://github.com/Pengfei-Y/INFSCI2415-FInalReport>

Google Colab Link: <https://colab.research.google.com/drive/1cMrAf7x1-UIUSoCDwVgVfazKmFPxcej?usp=sharing>