Emma Hoffmann

Analysis of California Real Estate Market: Factors Influencing Home Prices

Introduction

This report presents a comprehensive analysis of the real estate market for homes in California. The dataset under examination contains 120 observations and consists of 5 essential variables: state, price, size, number of bedrooms, and number of bathrooms. The purpose of this analysis is to explore how these variables influence home prices, with a particular focus on California. By answering specific research questions, we aim to gain insights into the real estate dynamics in the state.

The following research questions will be addressed in this report:

1. Using data only for California: How much does the size of a home influence its price?
2. Using data only for California: How does the number of bedrooms of a home influence its price?
3. Using data only for California: How does the number of bathrooms of a home influence its price?
4. Using data only for California: How do the size, the number of bedrooms, and the number of bathrooms of a home jointly influence its price?
5. Are there significant differences in home prices among the four states (CA, NY, NJ, PA)?

By addressing these questions, this report aims to provide valuable insights into the California real estate market and the factors affecting home prices in the state.

Data

The analysis involves using a dataset containing 120 observations, focusing on homes in California. The dataset comprises the following variables: State, Price, Size, Beds, and Baths.

Below is a brief overview of methodology steps:

- **Data Collection:** Data extracted from the accurate Lock5 sourced Excel file, which includes information concerning 120 observations.
- **Data Cleaning:** The dataset will undergo a data cleaning process, which includes handling missing values, addressing outliers, and ensuring data integrity.
- **Descriptive Analysis:** Descriptive statistical methods such as regression models and ANOVA tables will be used to provide a clear overview of the dataset.

Analysis

1. Using data only for California: How much does the size of a home influence its price?
A linear regression analysis was conducted using California-specific data to explore the relationship between home size and price. The analysis aimed to establish whether home size significantly impacts home prices in California.

The regression model produced the following key insights:

- **Regression Equation:** The regression equation generated by the analysis is: Price = -56.81675 + 0.33919 * Size. This equation indicates that the size of a home is positively associated with its

price. As the size of a home increases, the price is expected to increase. The coefficient for Size, 0.33919, represents the average increase in price for each one-unit increase in size.

- **Statistical Significance:** The coefficient for Size is statistically significant, with a p-value of 0.000463. This low p-value suggests that the relationship between size and price is not likely due to random chance. There is strong evidence to support the idea that the size of a home significantly influences its price in California.
- **R-squared Value:** The multiple R-squared value is 0.3594. This indicates that approximately 35.94% of the variation in home prices in California can be explained by the size of the home. While not all variation can be accounted for, this is a substantial portion, highlighting the importance of home size in determining prices.
- **F-statistic:** The F-statistic of 15.71 with a p-value of 0.0004634 demonstrates that the overall model is statistically significant. This suggests that the relationship between home size and price is meaningful and that the model provides valuable insights.

Interpretation: The analysis provides evidence that the size of a home *is* a significant factor influencing its price in California. For each additional unit of size, on average, a unit increase for the price of the home of 0.33919 can be anticipated. This is consistent with common real estate market dynamics, where larger homes tend to command higher prices.

```
Residuals:
   Min     1Q Median    3Q    Max
-462.55 -139.69  39.24 147.65 352.21


Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) -56.81675  154.68102  -0.367 0.716145
Size         0.33919    0.08558   3.963 0.000463 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 219.3 on 28 degrees of freedom
Multiple R-squared:  0.3594,     Adjusted R-squared:  0.3365
F-statistic: 15.71 on 1 and 28 DF,  p-value: 0.0004634
```

Figure 1: Linear Regression Analysis for Home Size vs. Price in CA

2. Using data only for California: How does the number of bedrooms of a home influence its price?

A linear regression analysis was conducted using California-specific data to examine how the number of bedrooms in a home influence its price. The analysis aimed to determine whether the number of bedrooms significantly impacts home prices in California.

The regression model produced the following key insights:

- **Regression Equation:** The regression equation generated by the analysis is: Price = 269.76 + 84.77 * Beds. This equation suggests that the number of bedrooms in a home is positively associated with its price. As the number of bedrooms increases, the price is expected to increase. However, the coefficient for Beds (84.77) indicates the average price increase for each additional bedroom.
- **Statistical Significance:** The coefficient for Beds is not statistically significant, with a p-value of 0.255. This higher p-value suggests that the relationship between the number of bedrooms and

price may not be statistically significant. There is limited evidence to support the idea that the number of bedrooms significantly influences home prices in California.

- **R-squared Value:** The multiple R-squared value is 0.04605, indicating that approximately 4.61% of the variation in home prices in California can be explained by the number of bedrooms. This R-squared value is relatively low, suggesting that the number of bedrooms alone does not explain much of the variability in home prices.
- **F-statistic:** The F-statistic of 1.352 with a p-value of 0.2548 indicates that the overall model may not be statistically significant. This suggests that the relationship between the number of bedrooms and home price may not be meaningful.

Interpretation: The analysis suggests that the number of bedrooms in a home in California *may not* have a statistically significant influence on its price. The coefficient for Beds is not statistically significant, and the low R-squared value indicates that the number of bedrooms alone does not explain much of the variability in home prices. This finding suggests that other factors may have a more significant impact on home prices in California, and the number of bedrooms may not be a strong predictor on its own.

```
Residuals:
    Min     1Q  Median     3Q     Max
-413.83 -236.62   29.94  197.69  570.94


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  269.76     233.62   1.155    0.258
Beds          84.77      72.91   1.163    0.255


Residual standard error: 267.6 on 28 degrees of freedom
Multiple R-squared:  0.04605,   Adjusted R-squared:  0.01198
F-statistic: 1.352 on 1 and 28 DF,  p-value: 0.2548
```
Figure 2: Linear Regression Analysis for Number of Bedrooms vs. Price in CA

3. Using data only for California: How does the number of bathrooms of a home influence its price?

A linear regression analysis was conducted using California-specific data to investigate how the number of bathrooms in a home influence its price. The objective was to determine whether the number of bathrooms significantly impacts home prices in California.

The regression model produced the following key insights:

- **Regression Equation:** The regression equation generated by the analysis is: Price = 90.71 + 194.74 * Baths. This equation indicates that the number of bathrooms in a home is positively associated with its price. As the number of bathrooms increases, the price is expected to increase. The coefficient for Baths (194.74) represents the average price increase for each additional bathroom.
- **Statistical Significance:** The coefficient for Baths is statistically significant, with a p-value of 0.00409. This low p-value suggests that the relationship between the number of bathrooms and price is unlikely to be due to random chance. There is strong evidence to support the idea that the number of bathrooms significantly influences home prices in California.
- **R-squared Value:** The multiple R-squared value is 0.2588, indicating that approximately 25.88% of the variation in home prices in California can be explained by the number of bathrooms. While not all variation can be accounted for, this is a substantial portion, highlighting the importance of the number of bathrooms in determining prices.

- **F-statistic:** The F-statistic of 9.779 with a p-value of 0.004092 demonstrates that the overall model is statistically significant. This suggests that the relationship between the number of bathrooms and home price is meaningful and that the model provides valuable insights.

Interpretation: The analysis provides strong evidence that the number of bathrooms in a home *significantly influences* its price in California. For each additional bathroom, there is an expected average price increase of 194.74 units. This finding aligns with common real estate market dynamics, where homes with more bathrooms tend to command higher prices.

The regression model explains approximately 25.88% of the variation in home prices, indicating that the number of bathrooms is a significant factor in determining home prices.

```
Residuals:
   Min     1Q Median    3Q    Max
-374.93 -181.56  -2.74  152.31  614.81


Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept)   90.71    148.57   0.611  0.54641
Baths        194.74     62.28   3.127  0.00409 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 235.8 on 28 degrees of freedom
Multiple R-squared:  0.2588,     Adjusted R-squared:  0.2324
F-statistic: 9.779 on 1 and 28 DF,  p-value: 0.004092
```
Figure 3: Linear Regression Analysis for Number of Bathrooms vs. Price in CA

4. Using data only for California: How do the size, the number of bedrooms, and the number of bathrooms of a home jointly influence its price?

In this analysis, a linear regression was performed using California-specific data to examine how the size of a home, the number of bedrooms, and the number of bathrooms jointly influence the home's price. The goal was to determine the combined impact of these three variables on home prices in California.

The regression model produced the following key insights:

- **Regression Equation:** The regression equation generated by the analysis is: Price = -41.5608 + 0.2811 * Size - 33.7036 * Beds + 83.9844 * Baths. This equation suggests that the size, number of bedrooms, and number of bathrooms jointly affect the home's price. Each of these variables is included in the equation, and their coefficients represent their respective influence on the price.
- **Statistical Significance:** The coefficient for Size is statistically significant with a p-value of 0.0259, indicating that the size of the home has a significant impact on the price. The coefficients for Beds and Baths are not statistically significant, with p-values of 0.6239 and 0.2839, respectively. This suggests that the number of bedrooms and bathrooms may not have a significant individual impact on home prices in California.
- **R-squared Value:** The multiple R-squared value is 0.3912, indicating that approximately 39.12% of the variation in home prices in California can be explained by the combination of size, number of bedrooms, and number of bathrooms. This is a moderate portion of the variation, emphasizing the importance of considering all three variables jointly.

- **F-statistic:** The F-statistic of 5.568 with a p-value of 0.004353 indicates that the overall model is statistically significant. This suggests that the combined influence of size, number of bedrooms, and number of bathrooms on home prices is meaningful.

Interpretation: The analysis reveals that when considering the size of a home, the number of bedrooms, and the number of bathrooms together, these variables *have a collective impact* on the home's price in California. The size of the home is statistically significant and positively associated with price, indicating that larger homes tend to have higher prices. However, the number of bedrooms and bathrooms, when considered jointly with size, do not show individual statistical significance in influencing home prices.

The regression model explains approximately 39.12% of the variation in home prices, suggesting that while these variables have an impact, there are other factors not considered in the model that also play a role in determining home prices in California.

```
Residuals:
   Min     1Q  Median    3Q     Max
-415.47 -130.32  19.64  154.79  384.94


Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept) -41.5608   210.3809  -0.198   0.8449
Size          0.2811     0.1189   2.364   0.0259 *
Beds        -33.7036    67.9255  -0.496   0.6239
Baths        83.9844    76.7530   1.094   0.2839
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 221.8 on 26 degrees of freedom
Multiple R-squared:  0.3912,      Adjusted R-squared:  0.3209
F-statistic: 5.568 on 3 and 26 DF,  p-value: 0.004353
```
Figure 4 Multiple Regression Analysis for Size, Bedrooms, and Bathrooms vs. Price in CA

5. Are there significant differences in home prices among the four states (CA, NY, NJ, PA)?
In this analysis, an analysis of variance (ANOVA) was performed to determine whether there are significant differences in home prices among the four states (California, New York, New Jersey, Pennsylvania). The objective was to assess whether the state in which a home is located has a significant impact on its price, using all available data.

The ANOVA test produced the following key insights:

- **ANOVA Table:** Degrees of Freedom (Df): The ANOVA table shows that there are three degrees of freedom for the State variable and 116 degrees of freedom for the residuals (unexplained variance).
- **Sum of Squares (Sum Sq):** The sum of squares for the State variable is 1,198,169, while the sum of squares for the residuals is 6,299,266.
- **Mean Square (Mean Sq):** The mean square is calculated by dividing the sum of squares by its respective degrees of freedom. For the State variable, the mean square is 399,390, and for the residuals, it is 54,304.
- **F-statistic:** The F-statistic is 7.355, and its associated p-value is 0.000148.

Emma Hoffmann

Interpretation: The results of the ANOVA test indicate that there are significant differences in home prices among the four states: California, New York, New Jersey, and Pennsylvania. The F-statistic of 7.355 with a p-value of 0.000148 suggests that the differences in home prices among these states are not likely due to random chance.

Therefore, it can be concluded that the state in which a home is located *has a statistically significant impact* on its price. The location, represented by the state, is a significant factor in determining home prices. This finding implies that real estate prices vary significantly depending on the state in which the property is situated, and additional factors associated with each state may play a role in influencing home prices.

```
         Df  Sum Sq Mean Sq F value   Pr(>F)
State      3 1198169  399390   7.355 0.000148 ***
Residuals 116 6299266   54304
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 5: ANOVA Results for Home Price Differences Among CA, NY, NJ, PA

Summary

- **Home Size vs. Price in California**: Linear regression shows a positive correlation, with home size accounting for approximately 36% of price variability.
- **Number of Bedrooms vs. Price in California**: The number of bedrooms appears to have a limited effect on price, contributing to only about 4.61% of its variability.
- **Number of Bathrooms vs. Price in California**: The number of bathrooms significantly influences the price, explaining around 25.88% of the price variation.
- **Combined Influence of Size, Bedrooms, and Bathrooms**: Together, these factors account for about 39.12% of price variability in California, with size being a significant individual factor.
- **Interstate Price Comparison**: Significant price differences are evident across CA, NY, NJ, PA, highlighting location as a key determinant in home pricing.

This report analyzes the California real estate market, focusing on the impact of size, bedrooms, and bathrooms on home prices from a dataset of 120 observations. Key findings indicate that home size and the number of bathrooms significantly affect home prices in California, while the number of bedrooms does not. Collectively, these factors explain about 39% of the pricing variation. Additionally, significant differences in home prices are evident across four states (CA, NY, NJ, PA), with location being a critical price determinant. Ultimately, this analysis underscores the importance of location in real estate valuation.

References

Anthropic. (n.d.). Sleep Study. Claude AI. https://claude.ai/

Lock, R. H., et al. (2021). Statistics: Unlocking the Power of Data, 3rd Edition. Datasets - Third Edition. SleepStudy. https://www.lock5stat.com/datapage3e.html

OpenAI. (n.d.). Sleep Study. ChatGPT. https://chat.openai.com/