WESTERN MICHIGAN UNIVERSITY

Analysis of Home Sale Prices in a Midwestern City: Factors Influencing Property Values

REPORT BY

SAMUEL JONES ARMOO

&

AUGUSTINE ENNIN

STAT 6620 APPLIED LINEAR MODELS

PROF. JOSHUA NARANJO

26TH NOVEMBER, 2024

TABLE OF CONTENTS

INTRODUCTION

STATEMENT OF THE PROBLEM

This project aims to analyze the sale price of homes in a midwestern city, using data collected from 522 individual home sales. The dataset provides a range of variables that could potentially influence home prices, including architectural style, construction quality, home size, number of bedrooms and bathrooms, presence of air conditioning, garage capacity, presence of a swimming pool, year of construction, lot size, and adjacency to highways. Each of these variables represents a different aspect of the home's features and can potentially impact its market value in different ways.

The primary objective of this analysis is to determine whether architectural style and construction quality have a statistically significant impact on the sales price of homes. By understanding the relationship between these features and home prices, this study aims to provide insights into the attributes that contribute most significantly to a home's value. Additionally, this study will identify other factors that significantly influence the sale price, such as home size, lot size, and presence of amenities like air conditioning and a pool. These insights will help to better understand how various attributes of a home affect its market value, either positively or negatively.

The analysis will employ statistical techniques such as Analysis of Variance (ANOVA) and multiple linear regression to explore differences in sales price across different architectural styles and construction qualities, as well as to quantify the impact of other contributing factors. ANOVA will be used to determine if there are statistically significant differences in mean sales prices among different categories of architectural styles and construction quality levels. Multiple linear regression will be employed to examine how various continuous and categorical variables jointly influence the sale price of homes. The final model will be achieved from model selection and refinement alongside some Automatic search procedures.

The results of this study are expected to provide useful information to potential homebuyers, real estate investors, and policymakers regarding the features that significantly affect home prices. For homebuyers, understanding which features add the most value can help in making informed purchasing decisions. For real estate investors, these insights can aid in identifying properties with the highest potential for value appreciation. Policymakers can also benefit from understanding market dynamics, which can inform urban planning and housing policy decisions aimed at enhancing affordability and housing quality.

RESEARCH QUESTIONS

This analysis seeks to answer the following key questions:

1. Architectural Style Influence: Is there a statistically significant difference in home sale prices among different architectural styles? Which styles are associated with higher sale prices?
2. Impact of Construction Quality: Does the quality of construction have a significant effect on sale prices? Do homes with high-quality construction fetch higher prices than those of medium or low quality?
3. Other Influential Factors: Beyond architectural style and construction quality, what other variables significantly affect sale prices? For each significant variable, what is the direction and magnitude of its influence?

OBJECTIVES

The objectives of this analysis are:

1. To determine if there is a significant difference in sale prices between different architectural styles.
2. To evaluate the impact of construction quality on home sale prices.
3. To identify other significant factors that influence home sale prices and quantify their effects.

DATA DESCRIPTION

The dataset used in this study consists of information collected from 522 individual home sales in a midwestern city. The variables included in the dataset provide a comprehensive overview of the features of each home that could potentially influence its sale price

Here is a list of the variables included in the dataset, along with their definitions and roles in the analysis:

1. ID: A unique identification number assigned to each home.
2. Sales Price: The sale price of the home, which is the target variable for this analysis.
3. Home Size: The finished area of the home, measured in square feet.
4. Bedrooms: The total number of bedrooms in the home.
5. Bathrooms: The total number of bathrooms in the home.
6. Air Conditioning: A binary variable indicating the presence of air conditioning (1 = Yes, 0 = No).
7. Garage: The number of cars that the garage can hold.
8. Pool: A binary variable indicating the presence of a swimming pool (1 = Yes, 0 = No).
9. Year of Construction: The year in which the home was originally constructed.
10. Quality: An index representing the quality of construction, categorized as high, medium, or low quality (1 = High, 2 = Medium, 3 = Low).
11. Style: The architectural style of the home, categorized into different styles.
12. Lot Size: The size of the lot on which the home is built, measured in square feet.
13. Highway Proximity: A binary variable indicating whether the home is adjacent to a highway (1 = Yes, 0 = No).

These variables will be analyzed to determine their impact on the sales price of homes. The focus will be on identifying the significant predictors and understanding how each variable influences the overall market value. The data contains both continuous variables (e.g., home size, lot size) and categorical variables (e.g., architectural style, construction quality) that will be used in the analysis to provide a holistic understanding of the factors affecting home prices

METHODOLOGY

DATA PREPARATION

- SAS® Studio was used for data entry, management and analysis. This process includes checking for missing values, ensuring data consistency, and preparing variables for analysis. Missing values, if any, will be sorted out to make analysis as smooth as possible.
- Categorical variables, such as architectural style, construction quality air conditioning, pool and highway proximity, will be encoded to allow their inclusion in the analysis.

DATA ASSESSMENT

- Descriptive statistics will be calculated to provide an overview of the dataset. This includes measures such as means, medians, standard deviations, and frequency distributions for categorical variables.
- Visualizations such as histograms, box plots, and scatter plots will be used to better understand the distribution of the data and identify any patterns or outliers. These visualizations will help in understanding the general trends and relationships among the variables.

ANALYSIS OF VARIANCE (ANOVA)

- One-way ANOVA was conducted to determine if there are statistically significant differences in mean sales prices among different architectural styles and construction quality levels. This analysis helped in identifying whether specific architectural styles or quality levels are associated with higher or lower sales prices.
- Post hoc tests, such as t Test LSD, was performed since significant differences were found, to determine which specific groups differed from each other.
- Two-Way ANOVA with interaction was performed to provide insights into how different combinations of architectural style and quality jointly impact the home sale price.

MULTIPLE LINEAR REGRESSION

- Multiple linear regression was employed to assess the relationship between sales price and various predictor variables. This model helped quantify the impact of each feature on the sales price while controlling for other factors.
- The regression model included both continuous variables and categorical variables. Interaction terms was considered to explore whether certain combinations of features have a joint effect on sales price.

MODEL SELECTION AND REFINEMENT

To develop a parsimonious model, several approaches were employed:

- R-squared, Adjusted and Akaike Information Criterion (AIC): R-squared, Adjusted R-squared selection and AIC criteria were used to evaluate the model fit and ensure that only meaningful predictors were retained.
- Stepwise Elimination: Variables were added step-by-step as in forward selection, but at each step, variables already in the model were also tested for significance and may be removed if they no longer contribute meaningfully. This ensured that only statistically significant variables were retained.

RESULTS AND DISCUSSION

The descriptive statistics provide an overview of the dataset, which includes 522 homes with variables. The average sale price was approximately $277,894, with a standard deviation of $137,923. This high standard deviation indicates a wide range of sale prices across different homes. The mean home size was 2,260 square feet, with considerable variability. Larger homes were generally more expensive, as expected. The mean lot size was 24,369 square feet, which also showed a wide range. Lot size was positively correlated with sale price, although not as strongly as home size (Table 1).

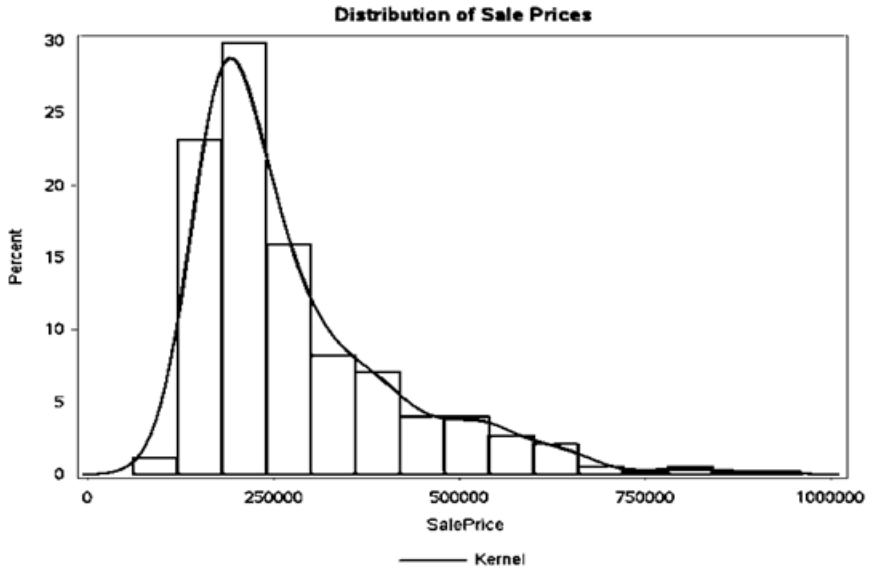
Table 1 Descriptive statistics for continuous variables

Variable	Mean	Std Dev	Minimum	Maximum
Sale Price	277894.15	137923.40	84000.00	920000.00
Home Size	2260.63	711.0659325	980.00	5032.00
Lot Size	24369.70	11684.08	4560.00	86830.00

For the frequencies of the categorical variables, 13.03% of homes are of Quality level 1(the highest quality). 55.56% of home fall under Quality level 2 whiles 31.42% of homes fall under Quality level 3. For Style, 41% of home are Style 1, making it the most common architectural style. 26.05% of home are Style 7, the next most common. Other styles are less frequent, with some styles like 10 and 11 appearing only once in the dataset. 83.14% of homes have air conditioning, while 16.86% do not. However, only 6.90% of homes have a pool, while the vast majority (93.10%) do not. 67.62% of homes have two-car garages, while 20.31% have three-car garages. There are few homes with up to seven-car garages or none at all. 97.89% of homes are not near a highway, while 2.11% are. The most common number of bedrooms was 2, with 38.70% of homes having two bedrooms. 34.29% of homes have 3 bedrooms, and 12.26% have 4 bedrooms. It was observed that, there are a small number of homes with 6 or more bedrooms, including one home with no bedrooms (Figure 1).

The distribution of sale prices showed positive skewness, indicating that while most homes were priced within a moderate range, a small number of homes had considerably higher prices. This suggests that there are more affordable homes compared to high-value properties, which might be due to market conditions, property features, or location factors (Figure 2).

Figure 2. Distribution of Home Sale Prices



ONE-WAY ANOVA FOR ARCHITECTURAL STYLE

The ANOVA table for architectural style shows that the F-value is 11.99 with a p-value < 0.0001. This indicates a statistically significant difference in the average sales price between the different architectural styles. Thus, we can reject the null hypothesis that there is no difference in sales prices among the architectural styles. The R-squared value is 0.1741, which indicates that around 17.41% of the variability in the Sale Price can be explained by the architectural style alone. This suggests that while style is a significant factor, other factors are likely contributing to the variability in home prices (Figure 3).

The LS-Means for each style indicate the average sales price after adjusting for multiple comparisons. Style 10 has the highest average sale price of $549,900, followed by Style 9 at $399,000. Conversely, Style 11 has the lowest average price at $159,900. Styles 6 and 7 also have higher LS-Means compared to the other styles, suggesting these styles tend to command a higher value (Table 2a).

The pairwise comparisons table shows which styles have statistically significant differences in their average sale prices. For instance: Style 10 and Style 7 is significantly different from styles 6, 5, 1, 3, 2, 4, and 11, indicating how they stand out as high-value styles. Style 11 consistently has significantly lower average sales compared to most other styles (Table 2b).

Table 2a Average Sale Price for each Architectural Style after adjusting for other factors.

Style	Sale Price LSMEANS	LSMEANS Number
1	248360.257	1
2	220022.310	2
3	248030.469	3
4	217568.182	4
5	251462.889	5
6	280252.778	6
7	369143.191	7
9	399000.000	8
10	549900.000	9
11	159900.000	10

Table 2b Significant Difference Between Architectural Types on Sale Price

Style Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		Significance
		Lower bound	Upper bound	
10-6	269647	14437	524857	***
10-5	298437	43227	553647	***
10-1	301540	52557	550523	***
10-3	301870	51533	552206	***
10-2	329878	79342	580413	***
10-4	332332	72883	591780	***
10-11	390000	38705	741295	***
7-6	88890	26587	151194	***
7-5	117680	55377	179984	***
7-1	120783	93542	148023	***
7-3	121113	83459	158767	***
7-2	149121	110165	188077	***
7-4	151575	73709	229441	***

ONE-WAY ANOVA FOR CONSTRUCTION QUALITY

The ANOVA results for Quality show an F-value of 503.87 with a p-value < 0.0001, indicating a statistically significant difference in sales prices across the different quality levels. The R-squared value is 0.6601, suggesting that around 66.01% of the variability in Sale Price can be explained by construction quality. This is a strong indication that quality is a major factor affecting home value (Figure 4).

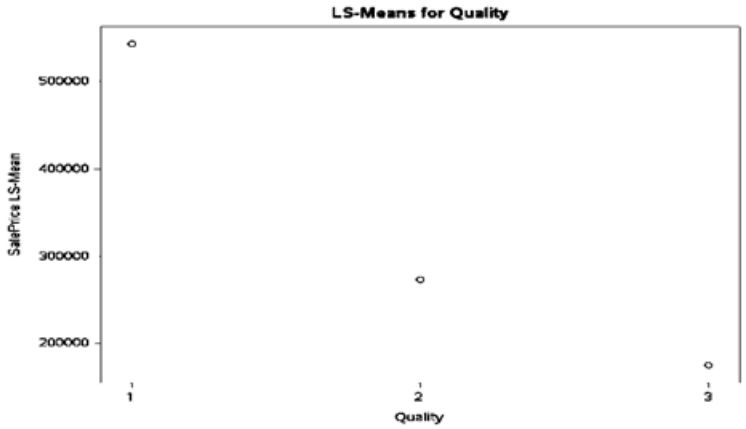
In Table 3a, the LS-Means indicate that homes with Quality level 1 have the highest average sales price at $543,610, followed by Quality level 2 at $273,766. Homes with Quality level 3 have the lowest average sales price at $175,018. This shows a clear trend where higher quality construction is associated with higher sales prices.

The LS-Means plot clearly shows a descending trend from Quality 1 to Quality 3. It further indicates the highest variability and median price for Quality 1, whereas Quality 3 has the lowest prices with less variability (Figure 5).

Table 3a Average Sale Price for each Construction Quality after adjusting for other factors.

Style	Sale Price LSMEANS	LSMEANS Number
1	543610.647	1
2	273766.279	2
3	175018.293	3

Figure 5. LS-Means for Quality Plot



The post-hoc comparisons indicate significant differences between all quality levels. The difference in sales price between Quality 1 and Quality 2 is $269,844, which is statistically significant. Similarly, there are significant differences between Quality 1 and Quality 3 ($368,592), and Quality 2 and Quality 3 ($98,748). These results confirm that higher quality homes consistently sell for more than lower quality ones (Table 3b).

Table 3b Significant Difference Between Quality Levels on Sale Price

Quality Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		Significance
		Lower bound	Upper bound	
1-2	269844	244328	295361	***
1-3	368592	341278	395907	***
2-1	-269844	-295361	-244328	***
2-3	98748	80245	117251	***
3-1	-368592	-395907	-341278	***
3-2	-98748	-117251	-80245	***

Comparisons significant at the 0.05 level are indicated by ***

TWO-WAY ANOVA FOR JOINT EFFECTS OF STYLE AND QUALITY ON SALE PRICE

The goal here was to determine whether combinations of Architectural Style and Quality of Construction jointly impact Sale Price. A Two-Way ANOVA was employed to help reveal if there's a statistically significant interaction between these factors. Results show that, the model had an overall F value of 59.62 with a p-value < 0.0001, indicating that the combination of architectural style and quality of construction significantly affects the sale price. The R-Square value of 0.7146 implies that approximately 71.46% of the variability in Sale Price can be explained by the model, which includes the individual and interaction effects of style and quality.

Looking at individual effect, architectural style had a significant effect on the sale price, which means that different styles have different average sale prices whereas quality of construction also had a strong effect on the sale price, with homes of higher quality commanding higher sale prices.

The interaction term (Style * Quality) was also statistically significant, indicating that the effect of one variable on sale price depends on the level of the other variable, indicating that these two variables are dependent of each other. This finding suggests that the impact of architectural style on sale price is influenced by the quality of construction, and vice versa (Figure 6).

Figure 6 Two-Way ANOVA for Joint Effects of Style and Quality on Sale Price

The GLM Procedure

Dependent Variable: SalePrice

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	21	7.0825759E12	337265519064	59.62	<.0001
Error	500	2.828336E12	5656671981.2		
Corrected Total	521	9.9109119E12			

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.714624	27.06457	75210.85	277894.1

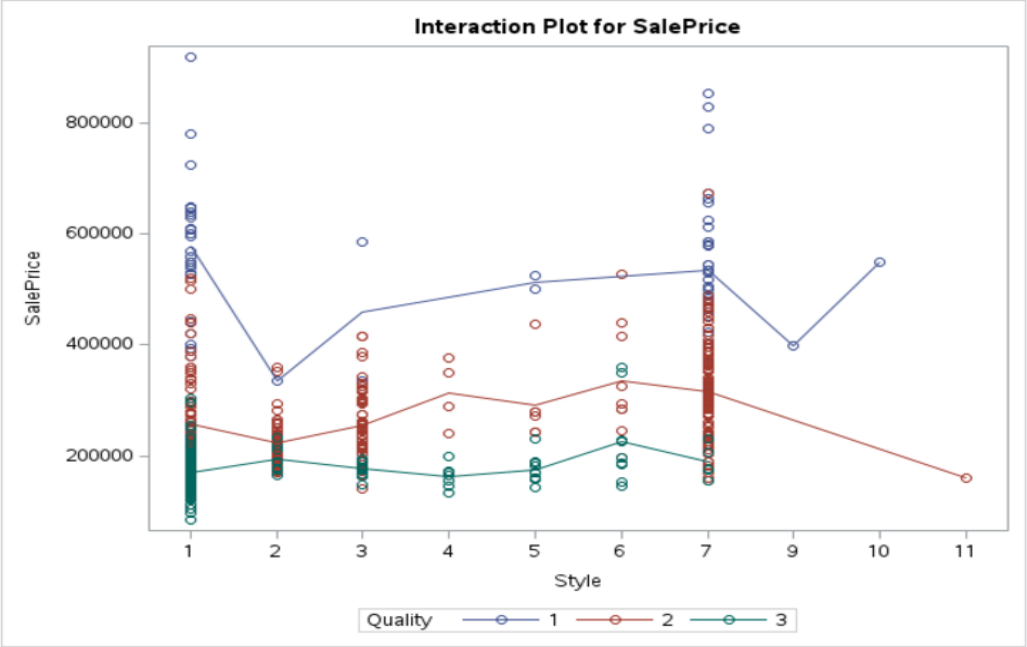
Source	DF	Type I SS	Mean Square	F Value	Pr > F
Style	9	1.7256618E12	191740201777	33.90	<.0001
Quality	2	5.1573521E12	2.578676E12	455.86	<.0001
Style*Quality	10	199562009761	19956200976	3.53	0.0002

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Style	9	164875063190	18319451466	3.24	0.0008
Quality	2	1.0495607E12	524780333564	92.77	<.0001
Style*Quality	10	199562009761	19956200976	3.53	0.0002

The interaction plot below provides a visual representation of how the combination of architectural style and quality affects the sale price. Holistically, homes with quality level 1 consistently show higher average sale prices across all architectural styles. Styles such as 7, 9, and 10 exhibit particularly high prices for quality level 1, indicating that these combinations are more desirable and fetch higher market values. Consequently, the lines for quality levels 2 and 3 are generally lower, indicating lower sale prices. However, some styles (such as style 7) have slightly higher prices compared to other styles, even at lower quality levels.

The interaction effect can be observed where the price trends for different styles vary across quality levels. For instance, styles 7, 9, and 10 have a steeper increase in sale price when moving from lower to higher quality compared to other styles. This interaction highlights the importance of considering both style and quality together when predicting sale prices (Figure 7).

Figure 7. Interaction Plot for Sale Price



MULTIPLE LINEAR REGRESSION MODEL WITH INTERACTION TERMS

The initial multiple linear regression model was extended by adding interaction terms between architectural style (Style) and quality of construction (Quality). This was done to evaluate whether certain combinations of these categorical variables jointly affected home sale prices

The model had an R-squared value of 0.847915, indicating that 84.79% of the variance in sale prices could be explained by the predictors in the model. This represented an improvement over previous models, suggesting that the interaction effects provided additional explanatory power (Figure 8).

It can be observed that each additional square foot of home size increases the sale price by approximately $109.57. Also, each additional square foot of lot size increases the sale price by $1.97. Each additional bathroom increases the sale price by $9,578.37. Construction increases sale price by $1,24.12 every year. Homes that are not near a highway are valued approximately $34,529 higher compared to those near a highway. However, Quality 1 (high quality) significantly increases sale price by approximately $144,325.32 compared to lower quality homes. Several architectural styles were found to have significant effects on sale price thus Style 1, Style 2, Style 3, and Style 4 were statistically significant predictors whiles Style 5, Style 6, Style 7, Style 9, Style 10, Style 11 were not.

The interaction between Quality and Style was added to assess joint effects. However, none of these interaction terms were found to be statistically significant at a 5% significance level. This implies that the effect of architectural style on the sale price is not significantly different based on the construction quality. In essence, the findings showed that Home Size, Lot Size, and Year were found to be the most influential continuous predictors with significant effects on sale price with quality of construction significantly contributes to higher sale prices, particularly for the best quality. Also, the interaction between Home Size and Lot Size was not significant, suggesting that the combined effect of larger home and lot sizes does not notably affect the sale price beyond their individual contributions.

Figure 8. Multiple Linear Regression with Interaction Terms – Initial Model

The GLM Procedure

Dependent Variable: SalePrice

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	31	8.4036131E12	271084293328	88.13	<.0001
Error	490	1.5072988E12	3076119995.5		
Corrected Total	521	9.9109119E12			

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.847915	19.95824	55462.78	277894.1

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	-2635141.986	B	404259.1621	-6.52	<.0001
HomeSize	109.577		11.3970	9.61	<.0001
LotSize	1.967		0.8020	2.45	0.0145
Bedrooms	-3783.137		3226.1004	-1.17	0.2415
Bathrooms	9578.370		4163.9628	2.30	0.0219
Garage	8481.598		4946.8036	1.71	0.0871
Year	1241.123		203.8895	6.09	<.0001
AirCon 0	-3380.318	B	7760.4869	-0.44	0.6633
AirCon 1	0.000	B	.	.	.
Pool 0	-11174.693	B	10128.4694	-1.10	0.2704
Pool 1	0.000	B	.	.	.
Highway 0	34529.160	B	17190.7239	2.01	0.0451
Highway 1	0.000	B	.	.	.
Quality 1	144325.315	B	27677.7723	5.21	<.0001
Quality 2	46163.864	B	24687.6887	1.87	0.0621
Quality 3	0.000	B	.	.	.
Style 1	127391.249	B	60756.7025	2.10	0.0365
Style 2	125369.577	B	63561.0743	1.97	0.0491
Style 3	129679.501	B	63082.9214	2.06	0.0403
Style 4	157094.834	B	64397.2793	2.44	0.0151
Style 5	120827.353	B	63381.3585	1.91	0.0572
Style 6	124325.864	B	63592.6222	1.96	0.0511
Style 7	62776.197	B	56575.0208	1.11	0.2677
Style 9	44514.866	B	79420.1351	0.56	0.5754
Style 10	40457.070	B	80919.7780	0.50	0.6173
Style 11	0.000	B	.	.	.
Quality*Style 1 1	52452.403	B	27980.7724	1.87	0.0614
Quality*Style 1 2	-105774.902	B	64146.0080	-1.65	0.0998
Quality*Style 1 3	-61979.562	B	50117.1684	-1.24	0.2168
Quality*Style 1 5	-78867.379	B	50899.4539	-1.55	0.1219
Quality*Style 1 7	0.000	B	.	.	.
Quality*Style 1 9	0.000	B	.	.	.
Quality*Style 1 10	0.000	B	.	.	.
Quality*Style 2 1	-40502.098	B	25479.8977	-1.59	0.1126
Quality*Style 2 2	-56954.715	B	31567.4056	-1.80	0.0718
Quality*Style 2 3	-47148.194	B	30336.4701	-1.55	0.1208
Quality*Style 2 4	-69719.168	B	43272.4133	-1.61	0.1078
Quality*Style 2 5	-60686.848	B	37692.0666	-1.61	0.1080
Quality*Style 2 6	-39086.827	B	35269.0824	-1.11	0.2683
Quality*Style 2 7	0.000	B	.	.	.
Quality*Style 2 11	0.000	B	.	.	.
Quality*Style 3 1	0.000	B	.	.	.
Quality*Style 3 2	0.000	B	.	.	.
Quality*Style 3 3	0.000	B	.	.	.
Quality*Style 3 4	0.000	B	.	.	.
Quality*Style 3 5	0.000	B	.	.	.
Quality*Style 3 6	0.000	B	.	.	.
Quality*Style 3 7	0.000	B	.	.	.
HomeSize*LotSize	-0.000		0.0003	-0.80	0.4232

REFINED REGRESSION MODEL

The adjusted regression model aimed to achieve a parsimonious representation of factors affecting home sale prices. The refined model focused on retaining only the statistically significant predictors and interactions, ensuring a balance between model complexity and interpretability. The model explains 84.60% of the variability in Sale Price (Figure 9). This is a strong indicator of the model's effectiveness in capturing most of the influential factors determining sale price. In this model, for each additional square foot of home size, the sale price increased

by $103.38. An increase in lot size by one square foot contributed approximately $1.34 to the sale price. However, newer homes were generally more valuable, with each additional year contributing $1,296.59 to the sale price. Each additional bathroom increases the sale price by $8,494.34. Highway proximity has a statistically significant effect on sale price of about &35,088.48. Also, higher quality construction (Quality Level 1) increased the sale price by approximately $149,409.52 compared to those with lower quality. Interestingly, Style 4 consistently showed a positive impact on the sale price by approximately $140,984.68 whiles the others were not.

The interaction between Style and Quality was significant, indicating that the effect of construction quality on sale price varied depending on the architectural style. For instance, Style 1 combined with Quality 1 significantly increased the sale price by $56,619.12.

Figure 9. Multiple Linear Regression with Interaction Terms – Refined Model

The GLM Procedure

Dependent Variable: SalePrice

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	26	8.3849895E12	322499596270	104.62	<.0001
Error	495	1.5259224E12	3082671490.7		
Corrected Total	521	9.9109119E12			

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.846036	19.97948	55521.81	277894.1

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	-2727992.125	B	392725.7199	-6.95	<.0001
HomeSize	103.380		7.1342	14.49	<.0001
LotSize	1.343		0.2273	5.91	<.0001
Bathrooms	8494.342		3938.8327	2.16	0.0315
Year	1296.593		197.6968	6.56	<.0001
Highway 0	35088.481	B	17181.5359	2.04	0.0417
Highway 1	0.000	B	.	.	.
Quality 1	149409.524	B	27282.6061	5.48	<.0001
Quality 2	47445.531	B	24409.8895	1.94	0.0525
Quality 3	0.000	B	.	.	.
Style 1	118728.934	B	60636.8216	1.96	0.0508
Style 2	117856.260	B	63315.5976	1.86	0.0633
Style 3	116993.701	B	62805.8981	1.86	0.0631
Style 4	140984.676	B	64041.5672	2.20	0.0282
Style 5	108459.037	B	63096.3647	1.72	0.0863
Style 6	114703.208	B	63337.6906	1.81	0.0708
Style 7	55515.182	B	56293.6682	0.99	0.3245
Style 9	33021.612	B	79290.8690	0.42	0.6773
Style 10	39753.243	B	80002.3006	0.50	0.6195
Style 11	0.000	B	.	.	.
Style*Quality 1 1	56619.122	B	27779.6741	2.04	0.0421
Style*Quality 1 2	-39088.334	B	25294.2471	-1.55	0.1229
Style*Quality 1 3	0.000	B	.	.	.
Style*Quality 2 1	-112109.912	B	64150.4327	-1.75	0.0812
Style*Quality 2 2	-56331.553	B	31498.8227	-1.79	0.0743
Style*Quality 2 3	0.000	B	.	.	.
Style*Quality 3 1	-46838.129	B	49593.9270	-0.94	0.3454
Style*Quality 3 2	-41933.550	B	30179.3228	-1.39	0.1653
Style*Quality 3 3	0.000	B	.	.	.
Style*Quality 4 2	-56823.560	B	42655.8709	-1.33	0.1834
Style*Quality 4 3	0.000	B	.	.	.
Style*Quality 5 1	-87063.012	B	49939.2290	-1.74	0.0819
Style*Quality 5 2	-59276.154	B	37420.2128	-1.58	0.1138
Style*Quality 5 3	0.000	B	.	.	.
Style*Quality 6 2	-37187.964	B	35263.2655	-1.05	0.2921
Style*Quality 6 3	0.000	B	.	.	.
Style*Quality 7 1	0.000	B	.	.	.
Style*Quality 7 2	0.000	B	.	.	.
Style*Quality 7 3	0.000	B	.	.	.
Style*Quality 9 1	0.000	B	.	.	.
Style*Quality 10 1	0.000	B	.	.	.
Style*Quality 11 2	0.000	B	.	.	.

MODEL SELECTION AND REFINEMENT USING STEPWISE SELECTION

The stepwise selection method was used to refine the model to achieve a more parsimonious version, while retaining all significant predictors that contribute meaningfully to the variation in Sale Price. The Akaike Information Criterion (AIC) was used as the selection criterion. The optimal value of the criterion was reached at Step 6, indicating that the following predictors should be included in the final model:

Table 4 Stepwise Selection Summary

Step	Effect Selected	Estimates	AIC
0	Intercept	-2680547	12880.1687
1	Style*Quality 1 1	277312	12267.6080
2	Home Size	103.380395	12025.2156
3	Year	1296.592556	11994.8218
4	Lot Size	1.343120	11960.9812
5	Bathrooms	8494.342406	11957.8670
6	Highway 0	35088	11955.4873*

*Optimal Value of Criterion

Model Statistics	
Root MSE	55522
Dependent Mean	277894
R-Square	0.8460
Adj R-Sq	0.8379
AIC	11955
AICC	11959
SBC	11546

From Table 4, results showed that there is no significant difference between the coefficients for the refined multiple linear regression model and that of the stepwise selection. However, Style 4 was not selected in this automatic search procedure. Here, each additional square foot of home size increases the sale price by $103.38. Each additional square foot of lot size adds $1.34 to the sale price. Also, additional bathroom increases the sale price by $8,494.32. Newer homes have a higher sale price, with each additional year increasing the value by $1,296.59. The proximity to highways significantly affects the sale price. In that, homes away from highways have a $35,088.41 higher price compared to those close to highways. The interaction effects indicate that the relationship between style and quality is important in determining the sale price. Thus, homes with Style 1 and Quality 1 have a significant effect on the sale price by $277,312 suggesting that certain styles, when combined with high quality, greatly increase the value of the property.

Finally, Home Size, Lot Size, Year of Construction, Bathrooms, and Highway Proximity are all significant individual predictors that influence the home sale price. However, the interaction effects indicate that the relationship between style and quality is important in determining the sale price. For instance, homes with Style 1 and Quality 1 have a significant effect on the sale price, suggesting that certain styles, when combined with high quality, greatly increase the value of the property.

The results from the stepwise selection indicate that the refined model retains significant predictors while removing variables that do not contribute meaningfully to the model, ensuring a parsimonious fit.

Therefore, the final model is:

Regression Equation:

$$\text{SalePrice} = -2680547 + 103.38 * \text{HomeSize} + 8494.34 * \text{Bathrooms} + 1296.59 * \text{Year} + 1.34 * \text{LotSize} + 35088 * \text{Highway\_0} + 277312.* (\text{Style\_1} * \text{Quality\_1})$$

CONCLUSION AND RECOMMENDATIONS

CONCLUSION

The analysis of home sale prices across various factors including home size, lot size, quality of construction, architectural style, and other property characteristics has yielded significant insights. Both architectural style and quality of construction have a significant effect on sale price. Homes with higher construction quality (Quality Level 1) command significantly higher prices, with styles 7, 9 and 10 also showing substantially higher desirability. The interaction between architectural style and quality significantly impacts the sale price. This means that the effect of quality differs depending on the architectural style, suggesting that high quality may be valued more in certain styles than others.

The findings from the multiple linear regression analysis with interaction terms and the subsequent model refinement indicate that a combination of home characteristics, including Home Size, Lot Size, Year of Construction, Quality, Style, and their interactions, significantly impact home sale prices. The interaction between Quality and Style was particularly notable, as it provided deeper insights into how the effect of construction quality is influenced by architectural style.

The model selection coupled with the stepwise selection procedure indicated that, Home Size, Lot Size and Year were found to be the most significant continuous predictors of home sale prices, with larger homes and lots commanding higher values. Properties away from the highway command a higher price compared to those near highways. However, homes with more bathrooms are valued higher, likely because additional bathrooms add convenience and value, particularly for larger households. Construction Quality significantly impacted sale prices, with higher quality homes fetching considerably higher values, specifically Quality Level 1. The interaction between Style and Quality was found to be statistically significant, emphasizing the importance of considering the joint effect of these features. For example, high-quality homes (Quality Level 1) in Style 1 showed significantly increased prices, while lower quality levels had reduced impact on sale prices across all styles.

These findings highlight the complexity of factors affecting home values and emphasize the need to consider both individual features and their joint effects when assessing property prices. The final model, which includes interaction terms, offers a comprehensive and delicate view of the determinants of home sale prices, explaining 84.60% of the variance in the dataset.

RECOMMENDATIONS

Developers and architects should focus on building homes with high-quality construction (Quality Level 1) and consider popular architectural styles such as styles 7 and 10, as these tend to yield higher sale

Real estate agents should market homes with higher construction quality and desirable styles as premium products. The interaction between style and quality should be leveraged to target buyers interested in maximizing value for their investments.

Future studies should consider incorporating more features, such as neighborhood characteristics, school quality, and proximity to amenities, which could help further explain the variance in sale prices. Additionally, efforts should be made to analyze influential observations that may skew the results, ensuring robustness in future models.

For older homes, enhancing the quality of construction and updating features to meet the characteristics of popular styles could significantly increase their market value. Renovations targeting these elements could yield a favorable return on investment.

APPENDIX

Figure 1. Frequency Distribution for Predictor Variables

Style	Frequency	Percent
1	214	41.00
2	58	11.11
3	64	12.26
4	11	2.11
5	18	3.45
6	18	3.45
7	136	26.05
9	1	0.19
10	1	0.19
11	1	0.19

Bathrooms	Frequency	Percent
0	1	0.19
1	71	13.60
2	171	32.76
3	175	33.52
4	84	16.09
5	17	3.26
6	1	0.19
7	2	0.38

Garage	Frequency	Percent
0	7	1.34
1	52	9.96
2	353	67.62
3	106	20.31
4	2	0.38
5	1	0.19
7	1	0.19

Bedrooms	Frequency	Percent
0	1	0.19
1	9	1.72
2	64	12.26
3	202	38.70
4	179	34.29
5	52	9.96
6	12	2.30
7	3	0.57

Quality	Frequency	Percent
1	68	13.03
2	290	55.56
3	164	31.42

Highway	Frequency	Percent
0	511	97.89
1	11	2.11

AirCon	Frequency	Percent
0	88	16.86
1	434	83.14

Pool	Frequency	Percent
0	486	93.10
1	36	6.90

Figure 3. ANOVA for Sale Price by Architectural Style

ANOVA for Sale Price by Architectural Style

The GLM Procedure

Dependent Variable: SalePrice

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	1.7256618E12	191740201777	11.99	<.0001
Error	512	8.1852501E12	15986816553		
Corrected Total	521	9.9109119E12			

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.174117	45.49897	126439.0	277894.1

Figure 4. ANOVA for Sale Price by Quality of Construction

ANOVA for Sale Price by Quality of Construction

The GLM Procedure

Dependent Variable: SalePrice

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	6.5417834E12	3.2708917E12	503.87	<.0001
Error	519	3.3691285E12	6491577138.9		
Corrected Total	521	9.9109119E12			

R-Square	Coeff Var	Root MSE	SalePrice Mean
0.660059	28.99317	80570.32	277894.1