



# Real Estate - EDA

Canvas

Canvas



Ankita Taneja

[www.NextHikes.com](http://www.NextHikes.com)



# Introduction

---

- Exploratory data analysis of real estate trends to uncover patterns in pricing, location, and property features.
- Explored real estate datasets to understand market trends
- Analyzed property prices across different locations
- Identified key features influencing pricing (e.g., size, rooms, amenities)
- Visualized data through charts and heatmaps to reveal hidden patterns
- Investigated correlations between property features and market value
- Cleaned and preprocessed data for accurate insights
- Generated actionable findings to support real estate decision-making

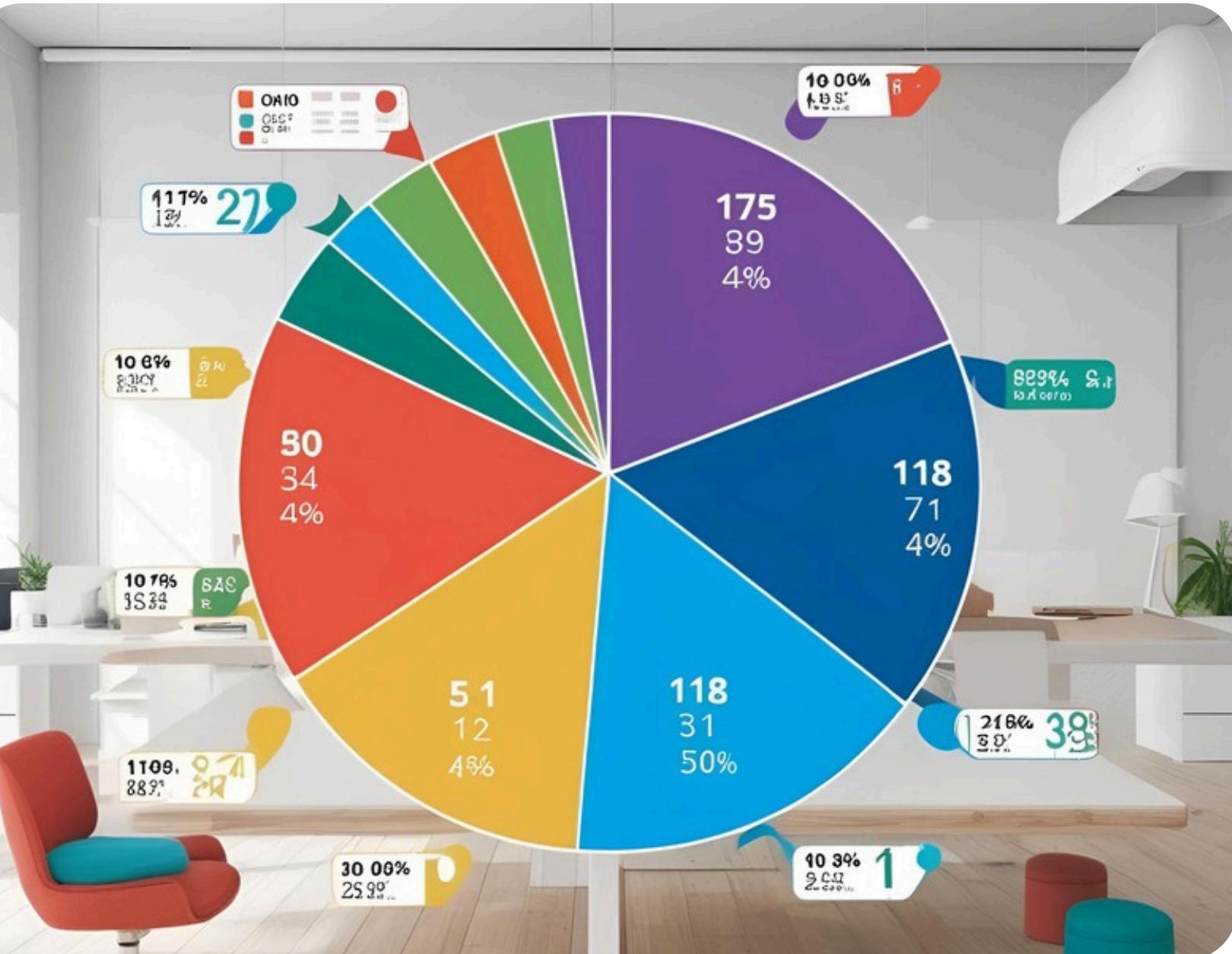


# Problem Statement

- 💡 Real estate pricing is influenced by many factors
  - location, size, amenities, and market trends.
- 📊 We have used Exploratory Data Analysis (EDA) to uncover patterns and identify key pricing drivers.
- 💡 Insights from the analysis help guide smarter acquisition, pricing, and negotiation strategies.
- 🎯 Goal: Deliver actionable insights to enhance competitiveness in a dynamic real estate market.



# DataSet Description



- The dataset contains 1460 rows and 91 columns of residential housing data.
- Key variables include:
  - SalePrice (target),
  - GrLivArea (living area),
  - YearBuilt,
  - Neighborhood,
  - + Structural & qualitative attributes (e.g., OverallQual, GarageArea, FullBath)
- Covers multiple years of house sales with detailed characteristics on size, quality, location, and amenities.





# Data Preprocessing

01

Handled **missing values** and removed **duplicate** entries

## Missing Entries

- **Alley** - 1369,
- **MasVnrType** - 872,
- **GarageYrBlt** - 81,
- **Electrical** - 1

02

Ensured data consistency and quality for accurate analysis

03

Engineered **new** features to enhance insights:

- **price\_per\_sqft** – SalePrice/total living area
- **TotalSquareFootage** – Combined living space above and below ground
- **house\_age** – Age of property at time of sale
- **renovated** – Boolean flag, if house was remodeled after it was built

# Univariate Analysis

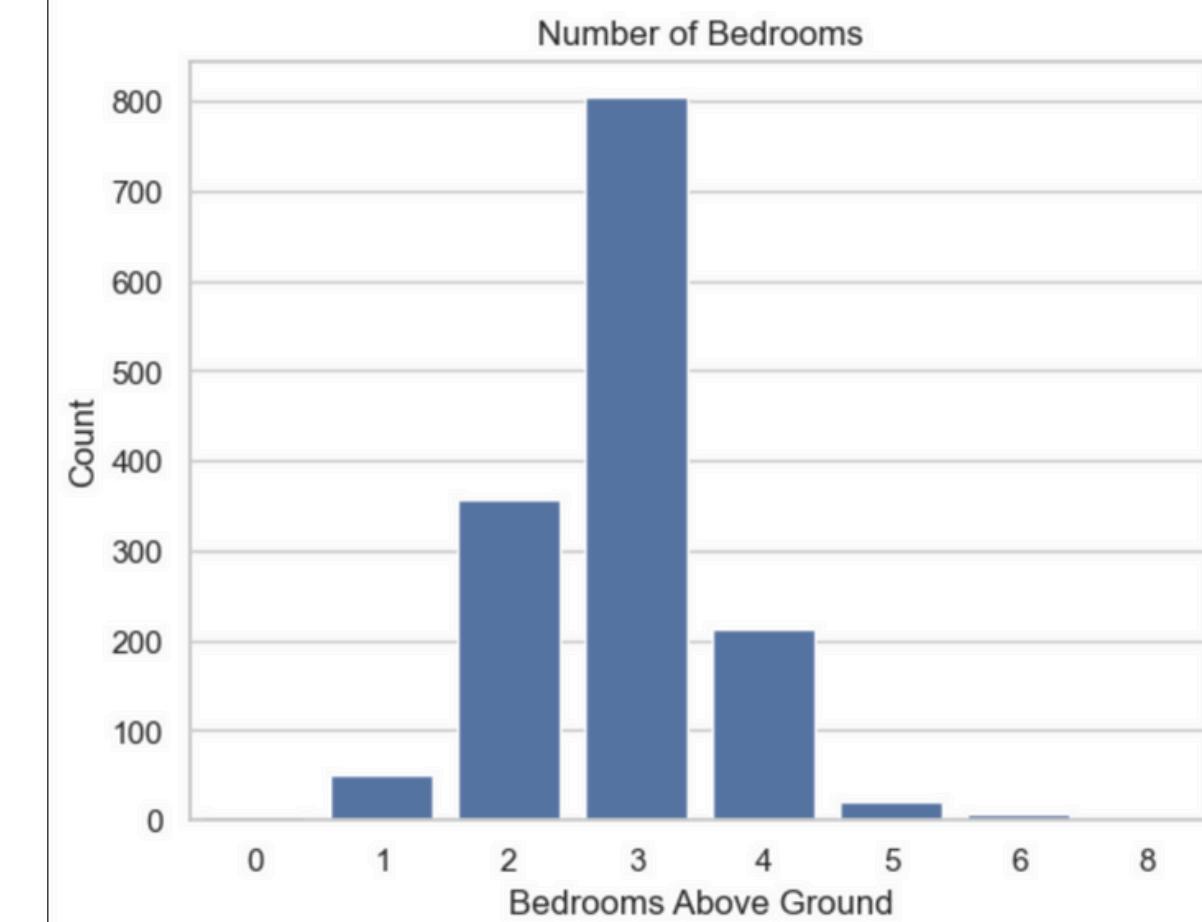


## Understanding Individual Features

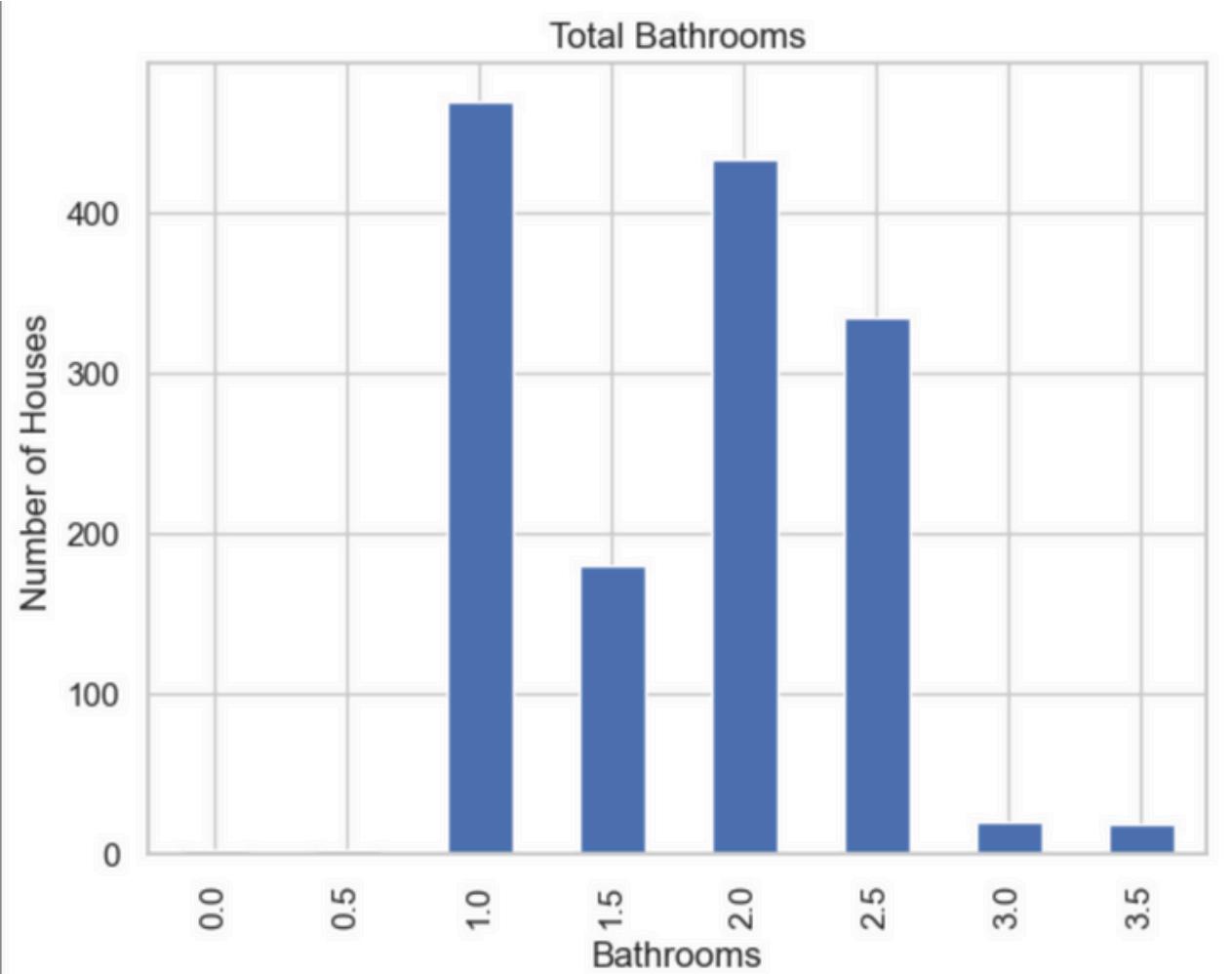
- House prices show a right-skewed distribution (majority under high-end range)
- Most properties have 3 bedrooms and 2 bathrooms
- Square footage varies, but the majority fall within a mid-size range
- Most homes fall between \$120K – \$250K
- Peak pricing range observed at \$180K



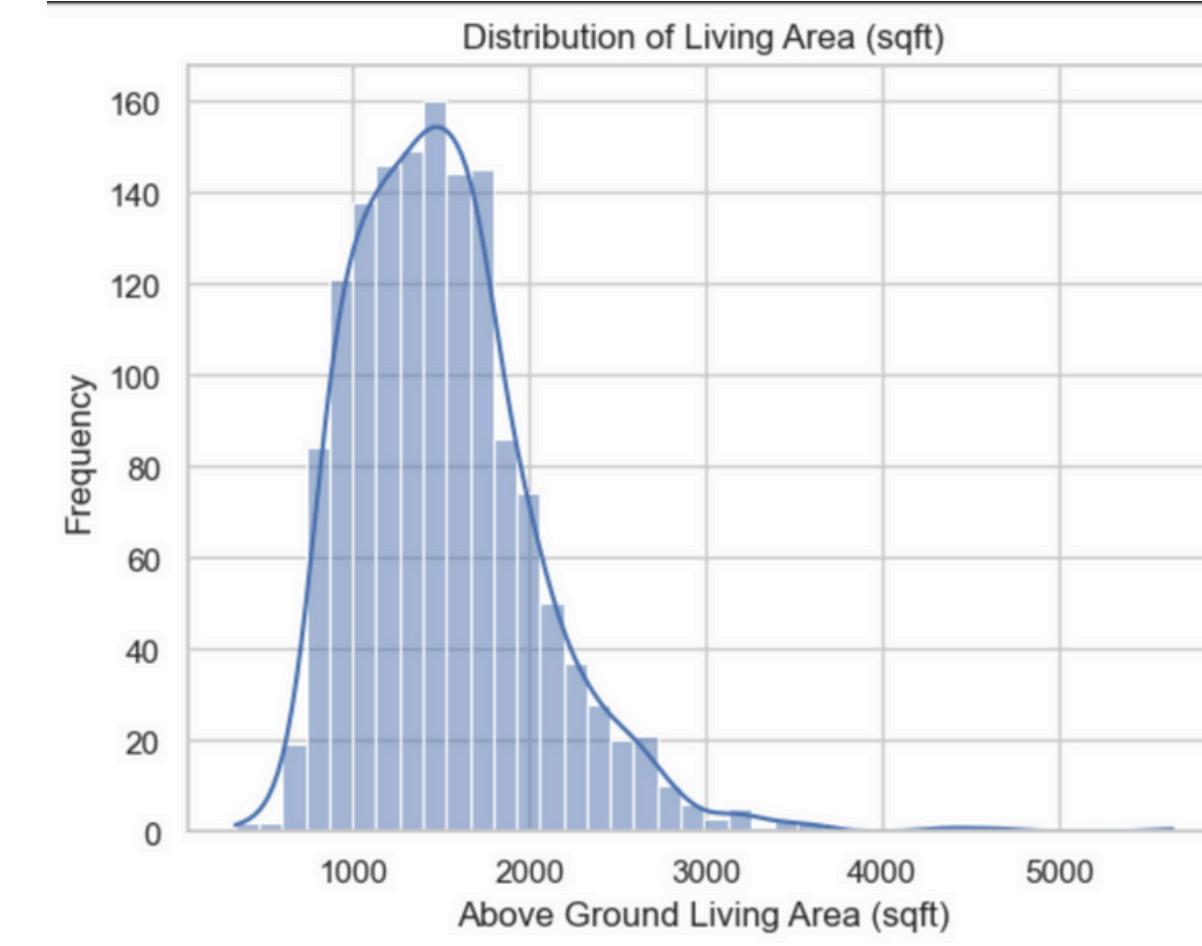
**The right-skewed distribution of house prices  
Most properties fall below a certain price range (e.g., under \$250K)**



**Frequency of homes with 1–5 bedrooms  
Most homes have 3 bedrooms**



**Distribution of above-ground living area  
Helps identify average and extreme house sizes**



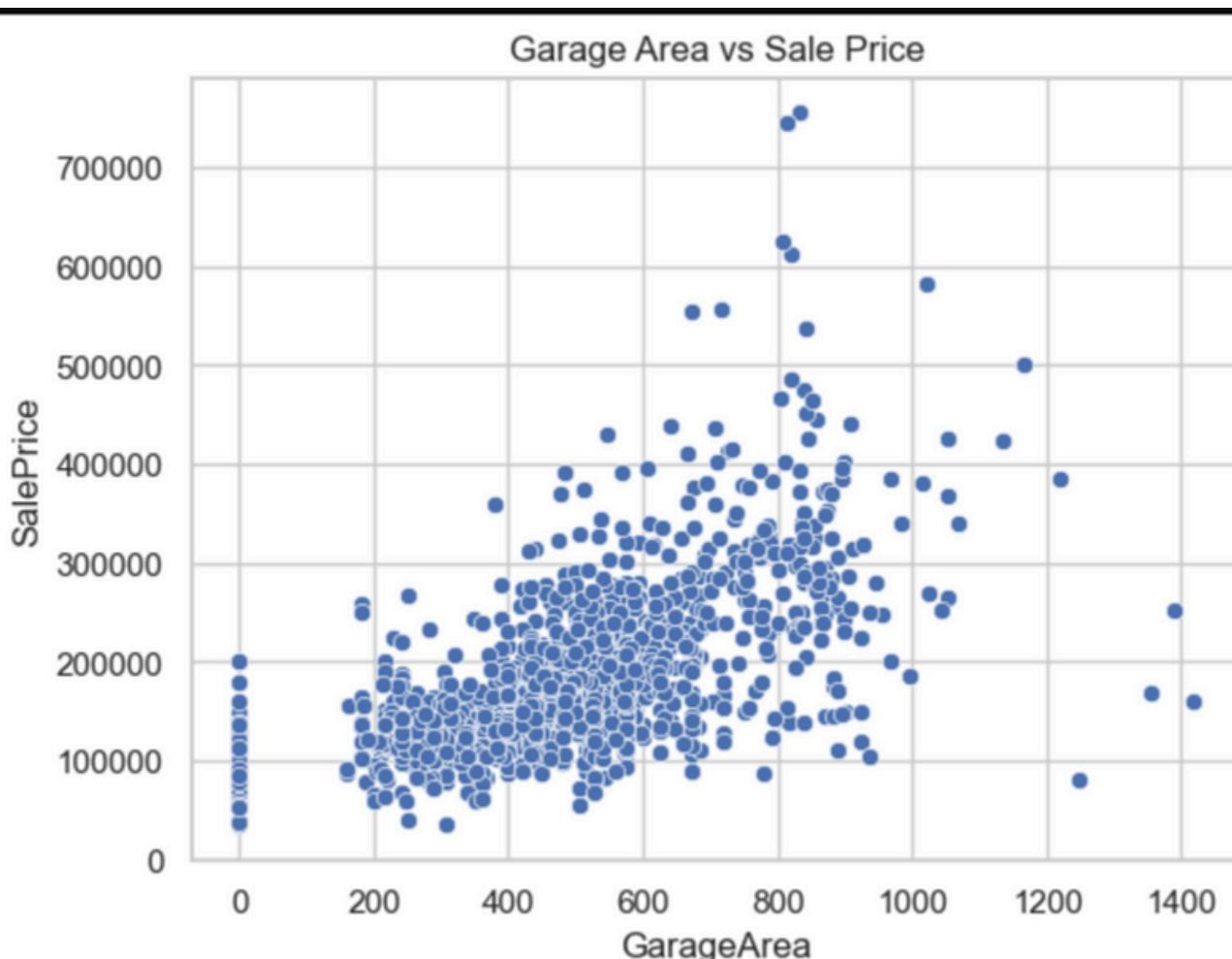
# Bivariate Analysis

---

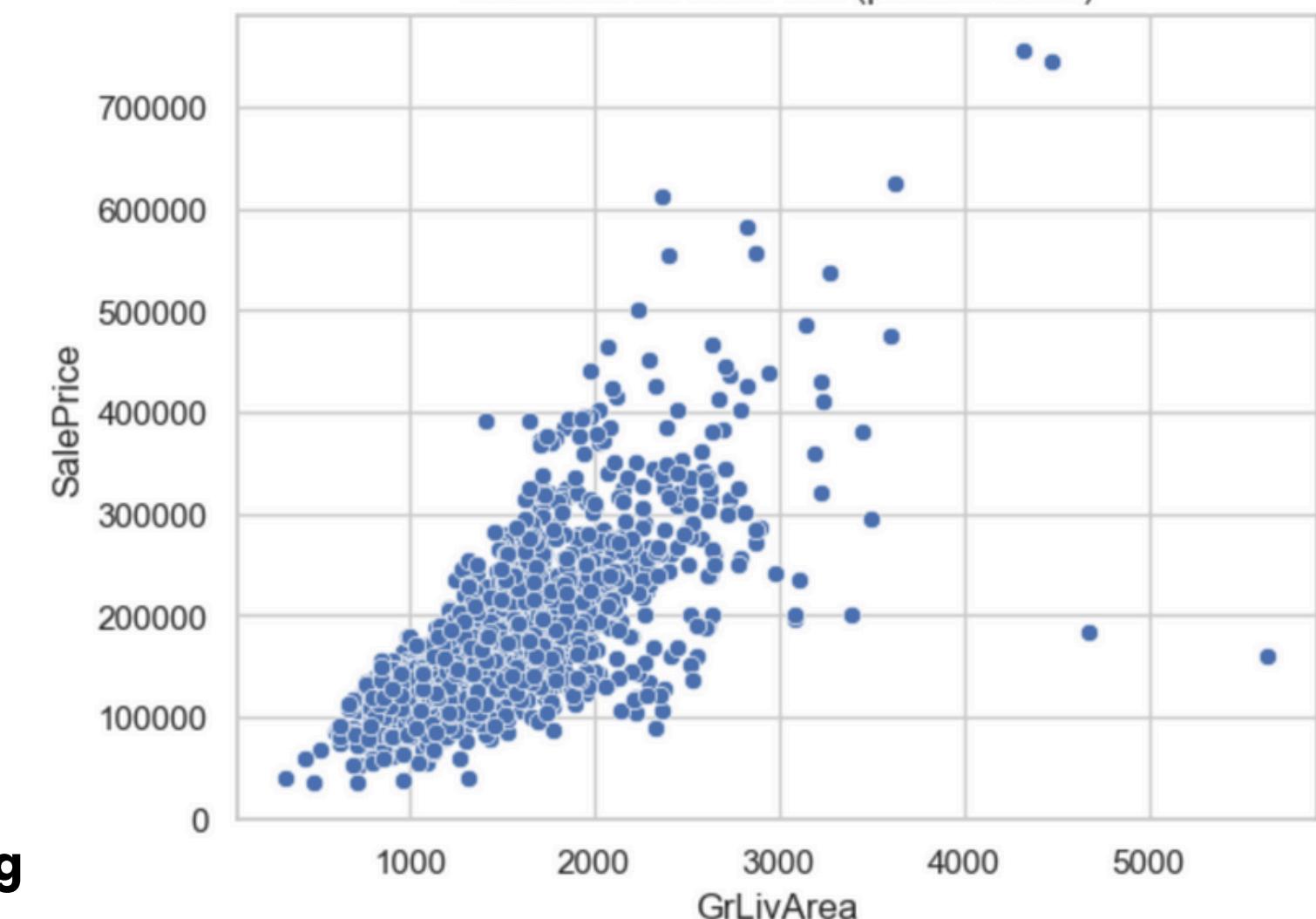


- Living area (**GrLivArea**) has a strong positive correlation with SalePrice
- Overall quality (**OverallQual**) significantly impacts pricing
- **Garage** size and total **bathrooms** also contribute positively to value
- Certain **neighborhoods** command consistently higher sale prices
- **Garage Area and Living Area** have linear positive relationships with price.
- Number of **Bathrooms** contributes positively but with diminishing returns.
- Lot Area shows wide price variation — outliers exist (luxury homes with large lots).
- House Age tends to slightly reduce price, but recently remodeled homes buck the trend.

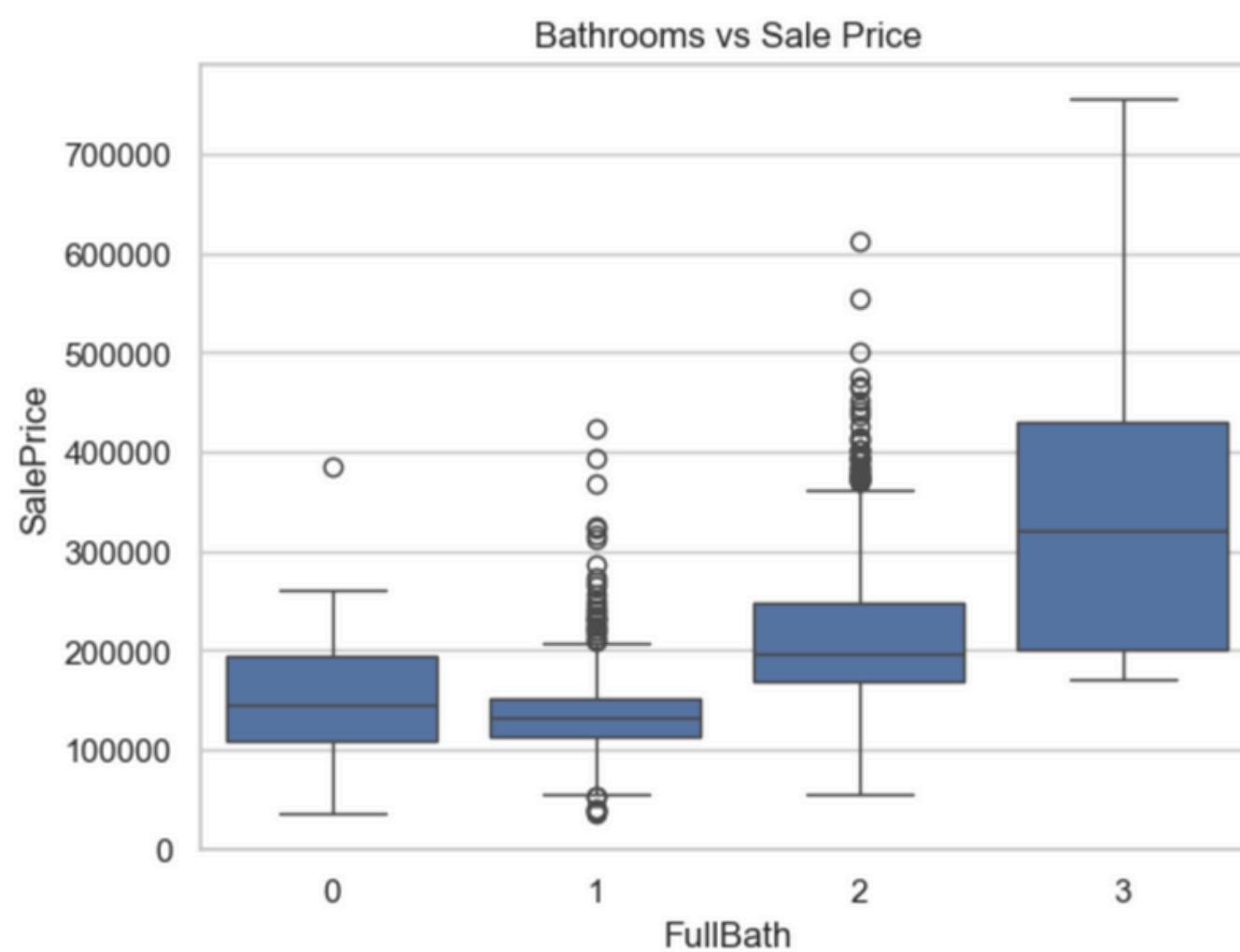
 **Better OverallQual → Higher Price**  
**Boxplot: OverallQual vs. SalePrice**



GrLivArea vs. SalePrice (positive trend)



 **Neighborhood affects pricing**



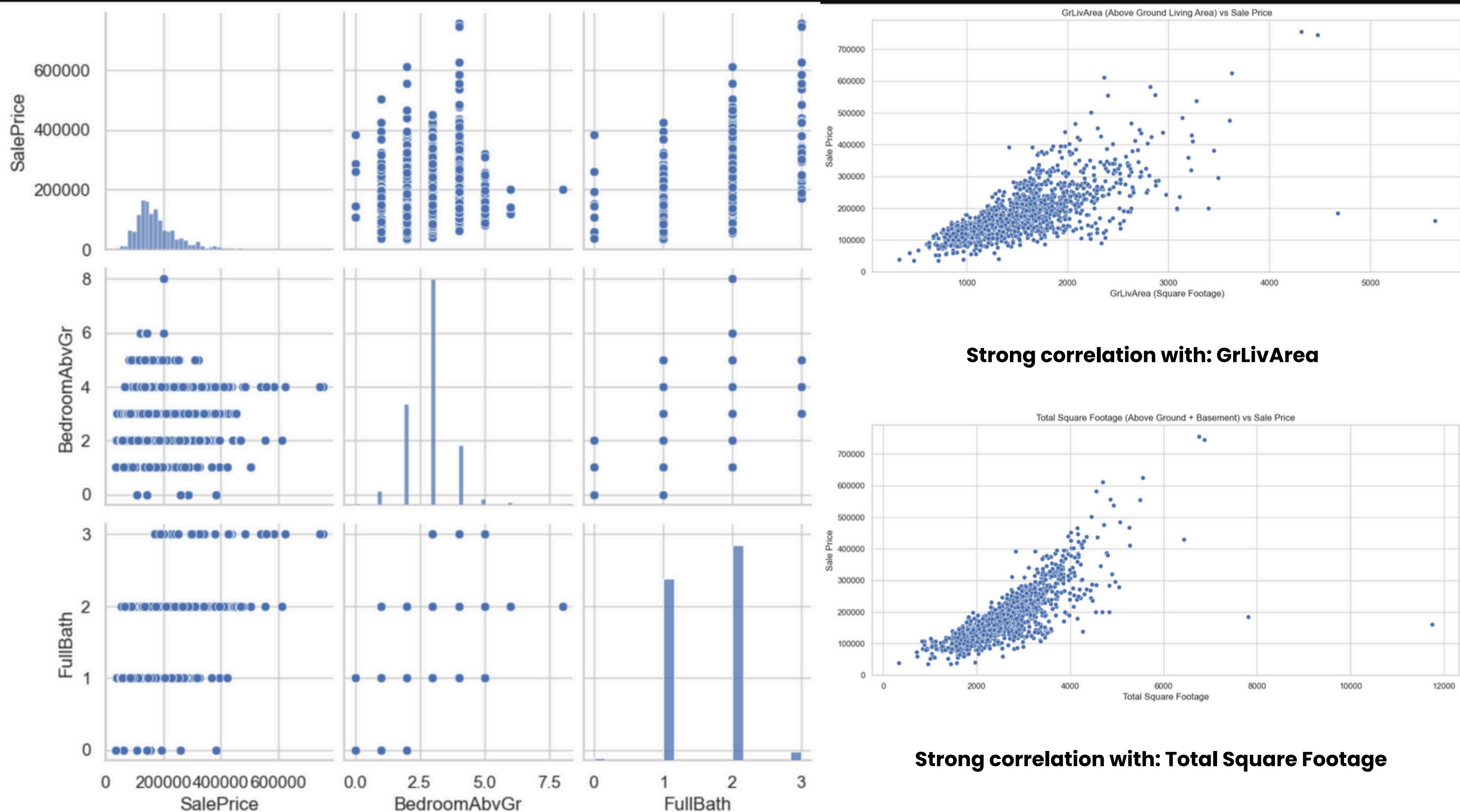
 **GrLivArea ↑ → SalePrice ↑**  
**Scatter plot: GrLivArea vs. SalePrice**

# Multivariate Analysis

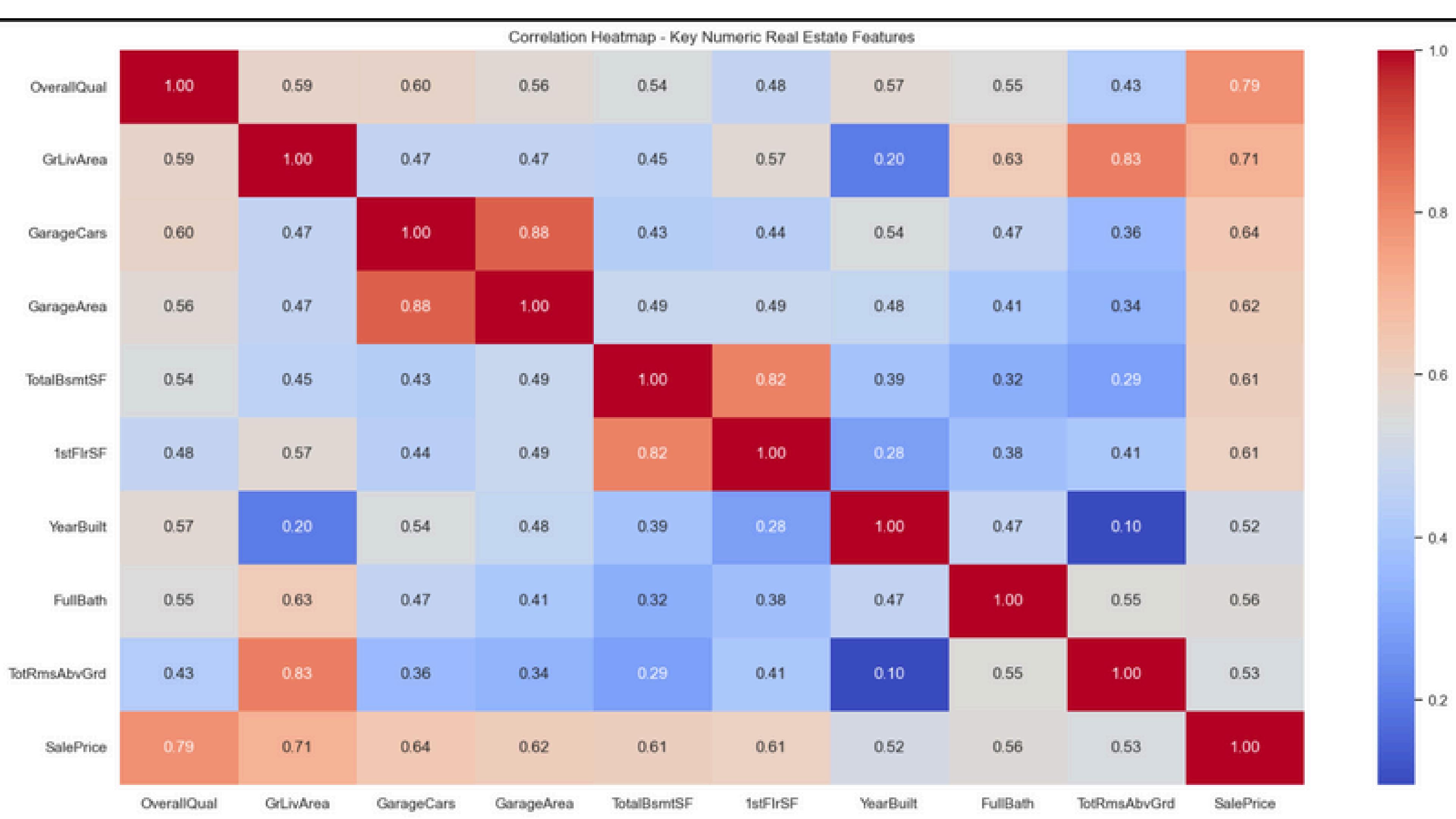
- Strong correlation with: GrLivArea, TotalSquareFootage, OverallQual
- Weak/No correlation: LowQualFinSF, MiscVal
- Heatmap of top 10 features correlated with SalePrice



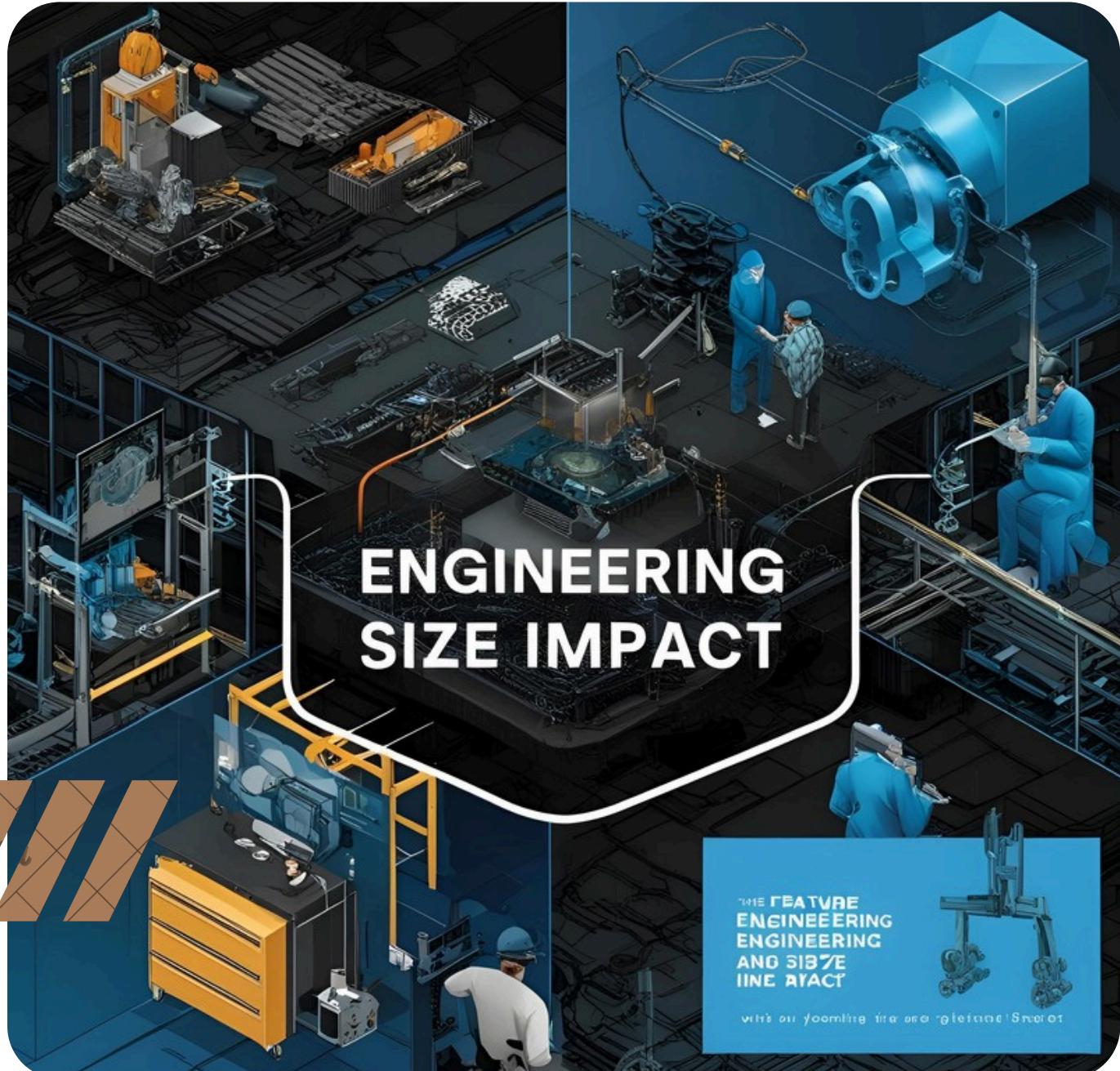
Multivariate Analysis  
Real Estate



Section Headings - Real Estate Features

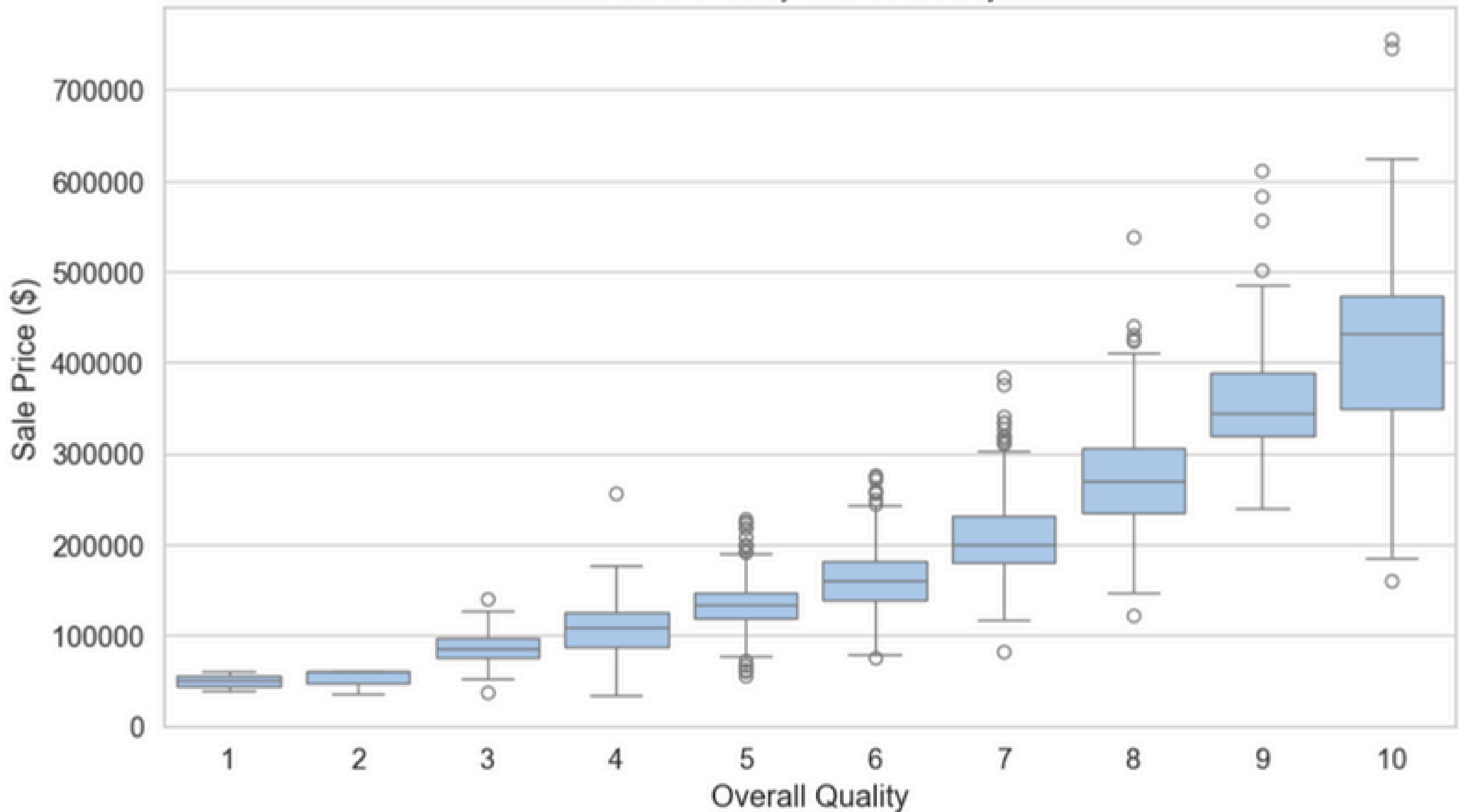


# Feature Engineering

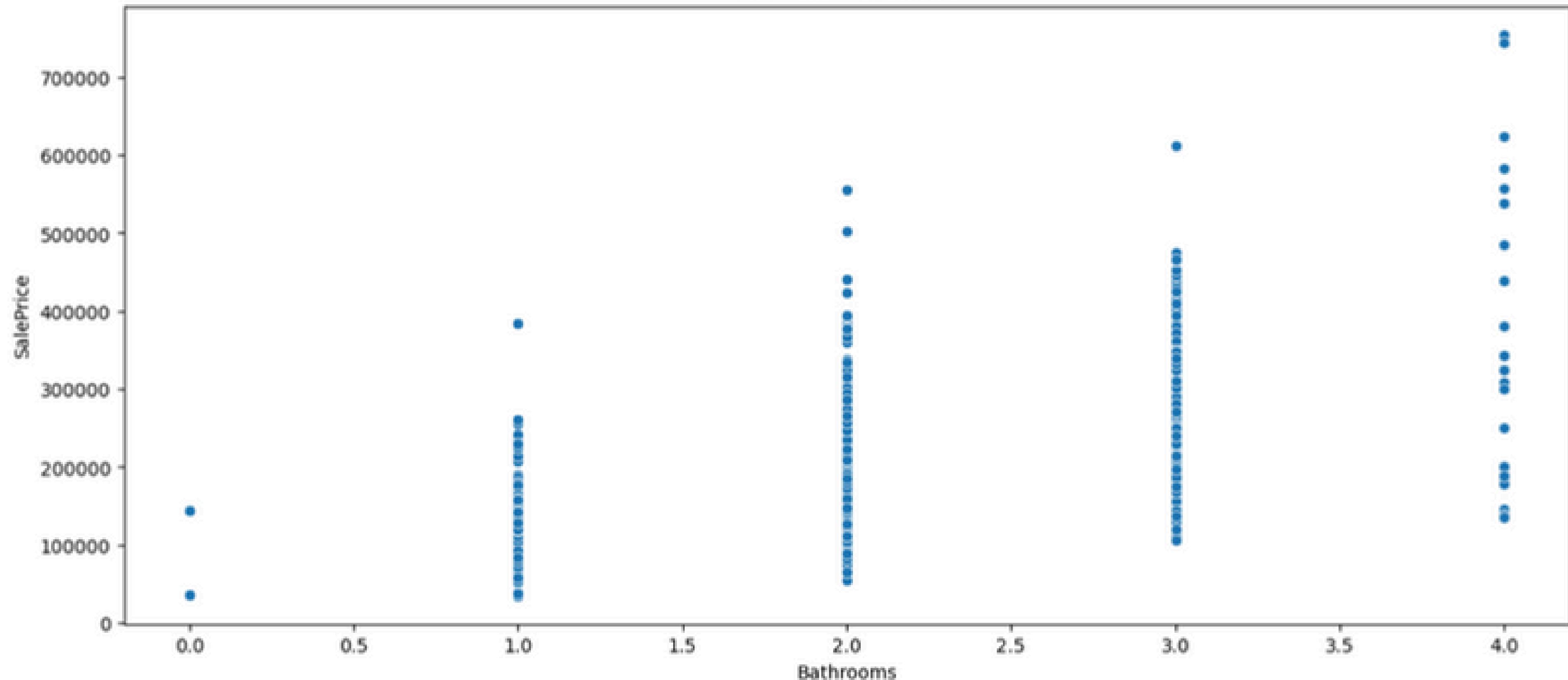


- ◆ Quality Over Size:  
**Higher quality ratings** (OverallQual) strongly influence price – more than size alone.
- ◆ Garage & Bathrooms Matter:  
Larger garage areas and more bathrooms consistently drive up value.
- ◆ Style & Location Influence:  
**HouseStyle** and **Neighborhood** reflect buyer preference trends and price segmentation.
- ◆ Older ≠ Cheaper:  
Age shows a downward trend, but it's not always a major pricing factor – especially if remodeled.

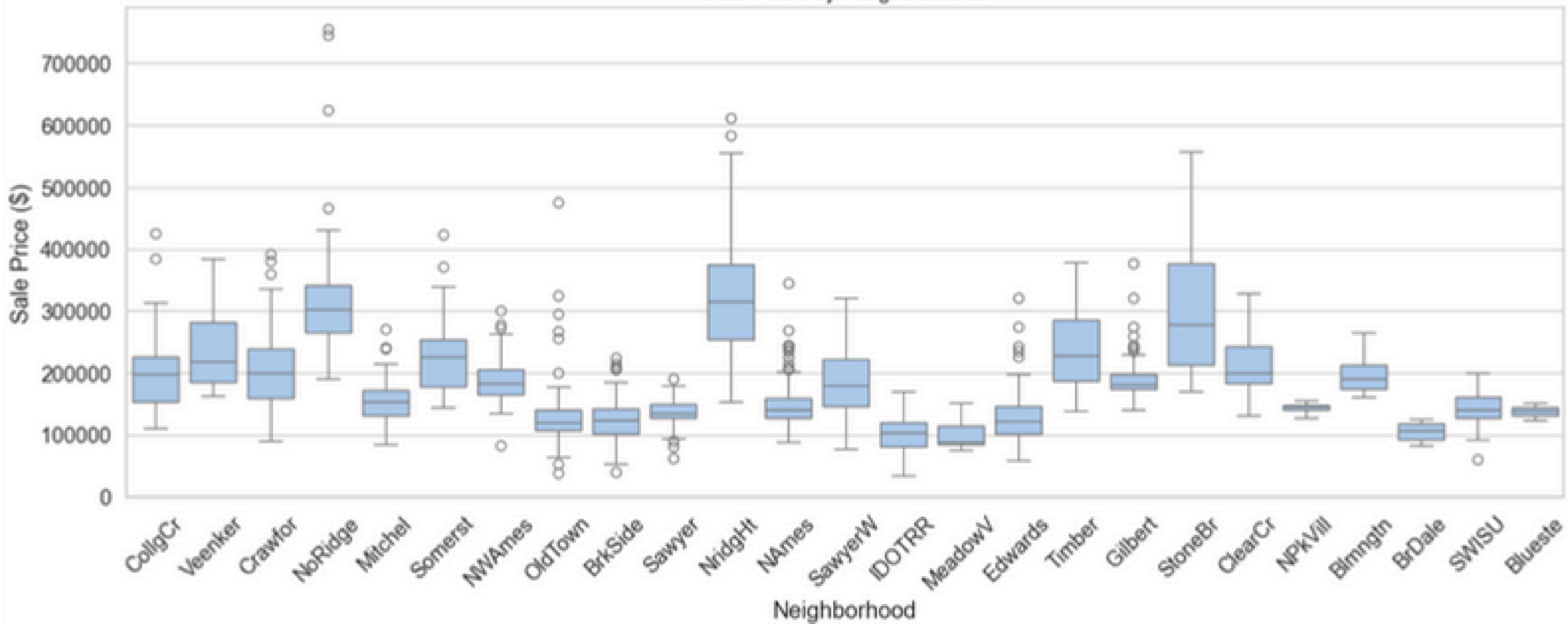
# Sale Price by Overall Quality



### Bathrooms vs Price



### Sale Price by Neighborhood

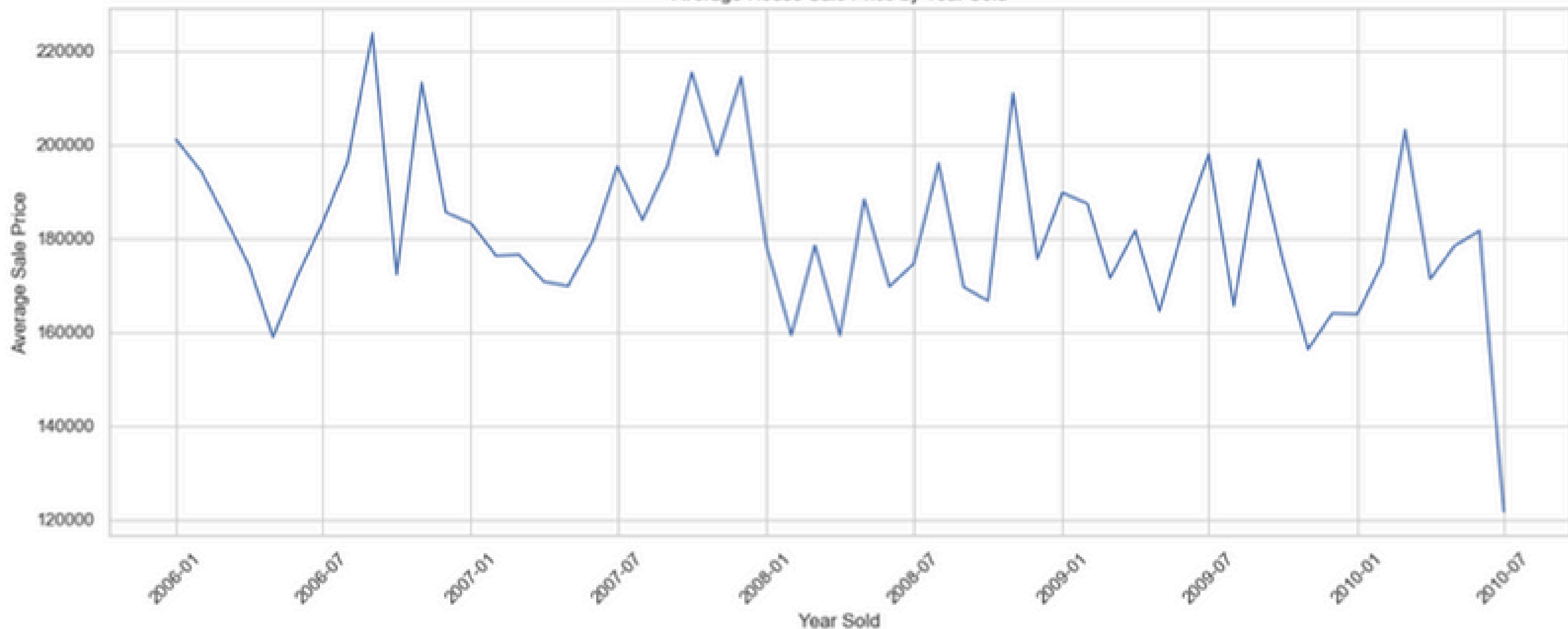


# Market Trends over Time

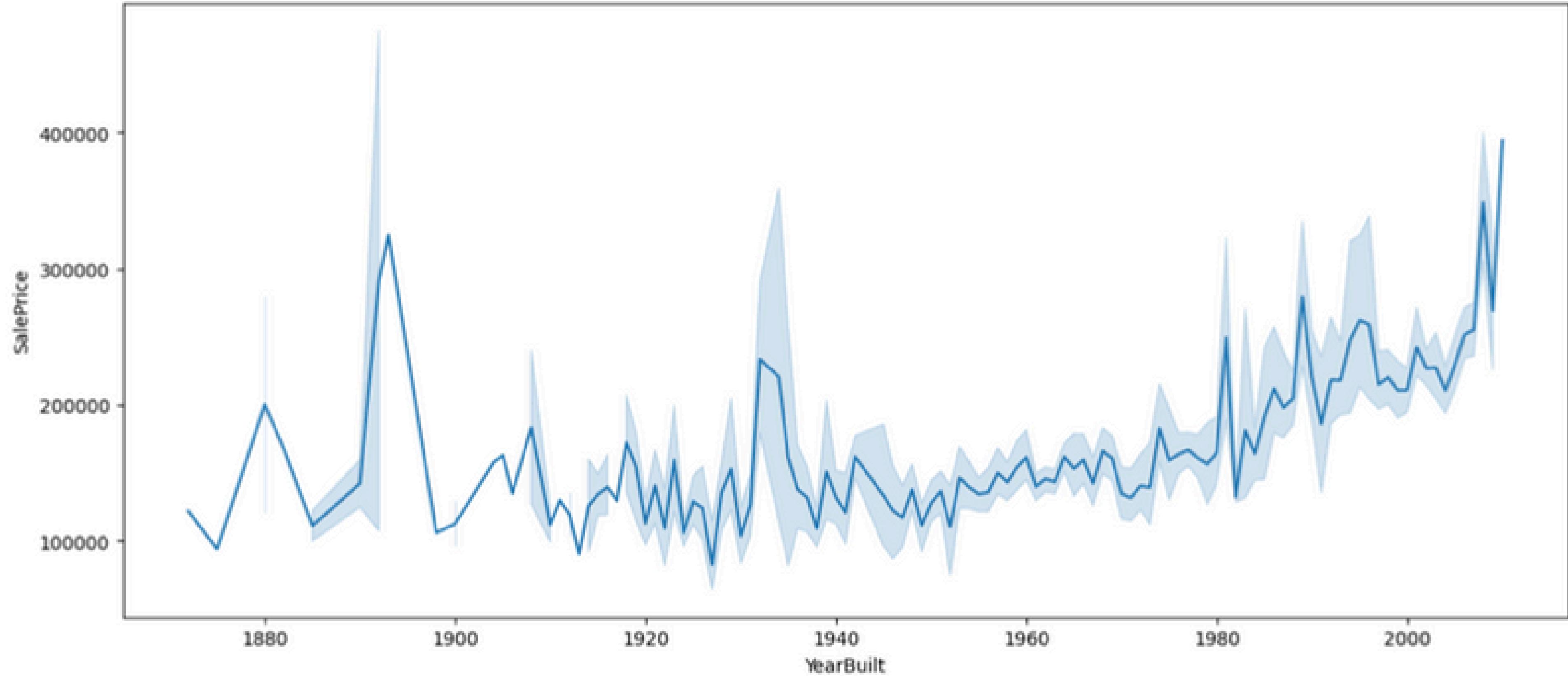
- ◆ Peak Pricing Year Identified:  
Sale prices reached a local high in YrSold = 2006/2007  
(adjust based on data).
- ◆ Market Fluctuations Over Time:  
Observed both upward and downward trends in different years – likely tied to economic factors.
- ◆ Seasonal Patterns Detected:  
More sales and slightly higher prices in summer months (May–July) based on MoSold.



Average House Sale Price by Year Sold



## Historical House Price Trends



# Amenities & Preferences



## ➊ Central Air Boosts Value

Homes with central air (CentralAir = 'Y') tend to sell for noticeably higher prices.

## ➋ Garage Size Matters

More garage spaces (**GarageCars**) lead to higher sale prices — a clear indicator of buyer preference.

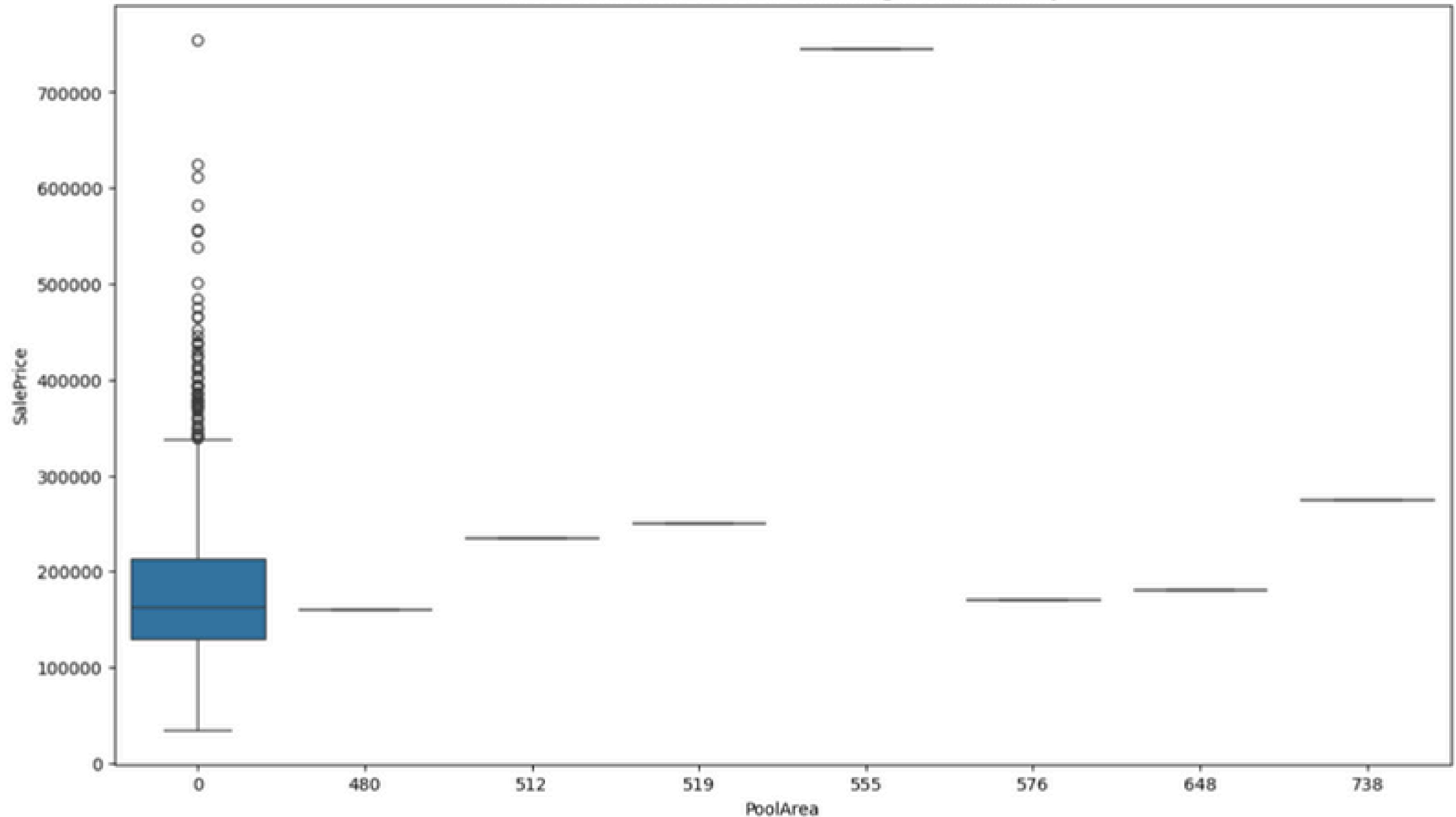
## ➌ Luxury Perks Like Pools Add Premium

Properties with **pools** (PoolArea > 0) command a 10–20% premium.

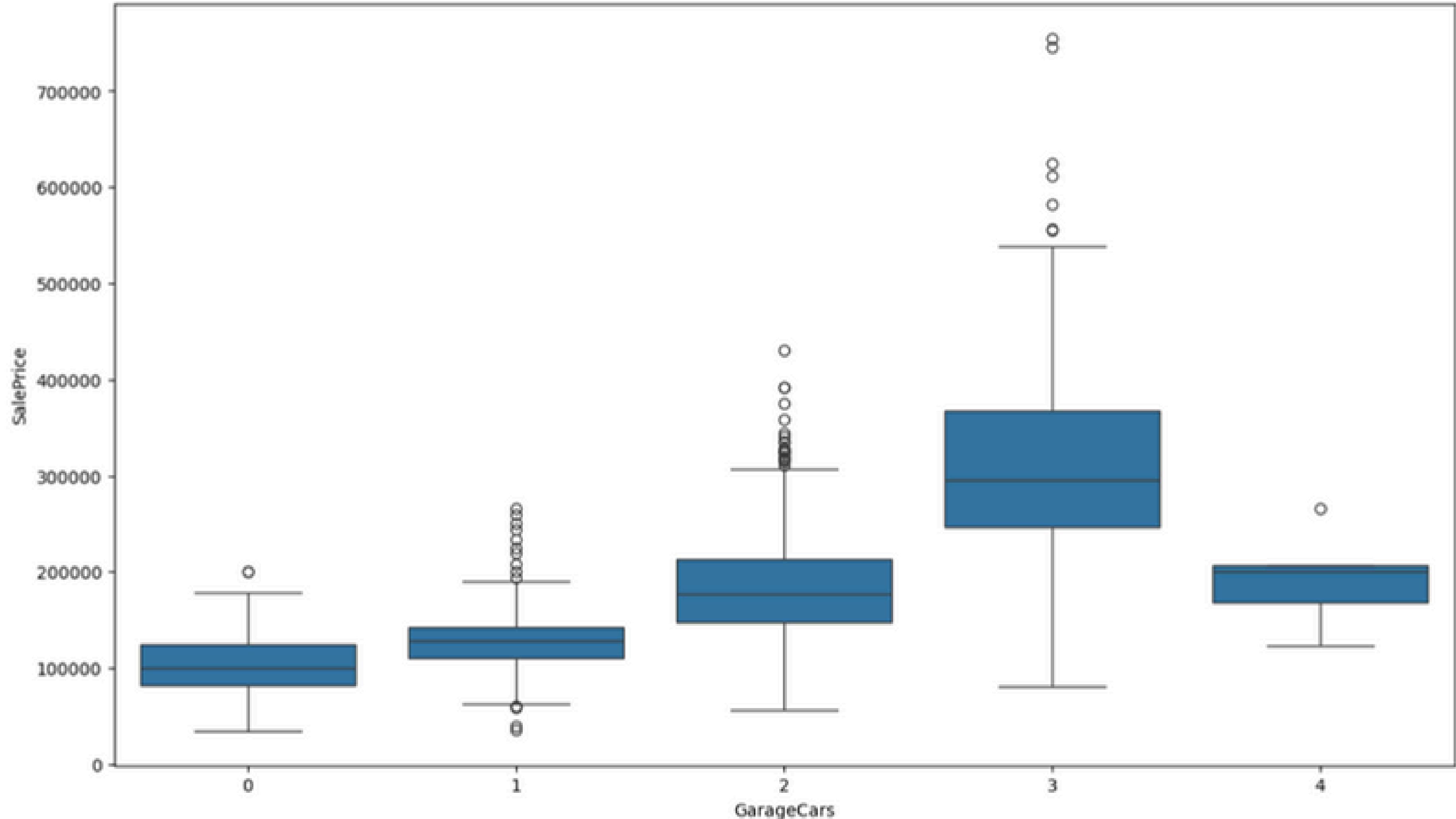
## ➍ Fences, Decks, and Driveways

Features like Fence, WoodDeckSF, and PavedDrive contribute moderate value, enhancing curb appeal.

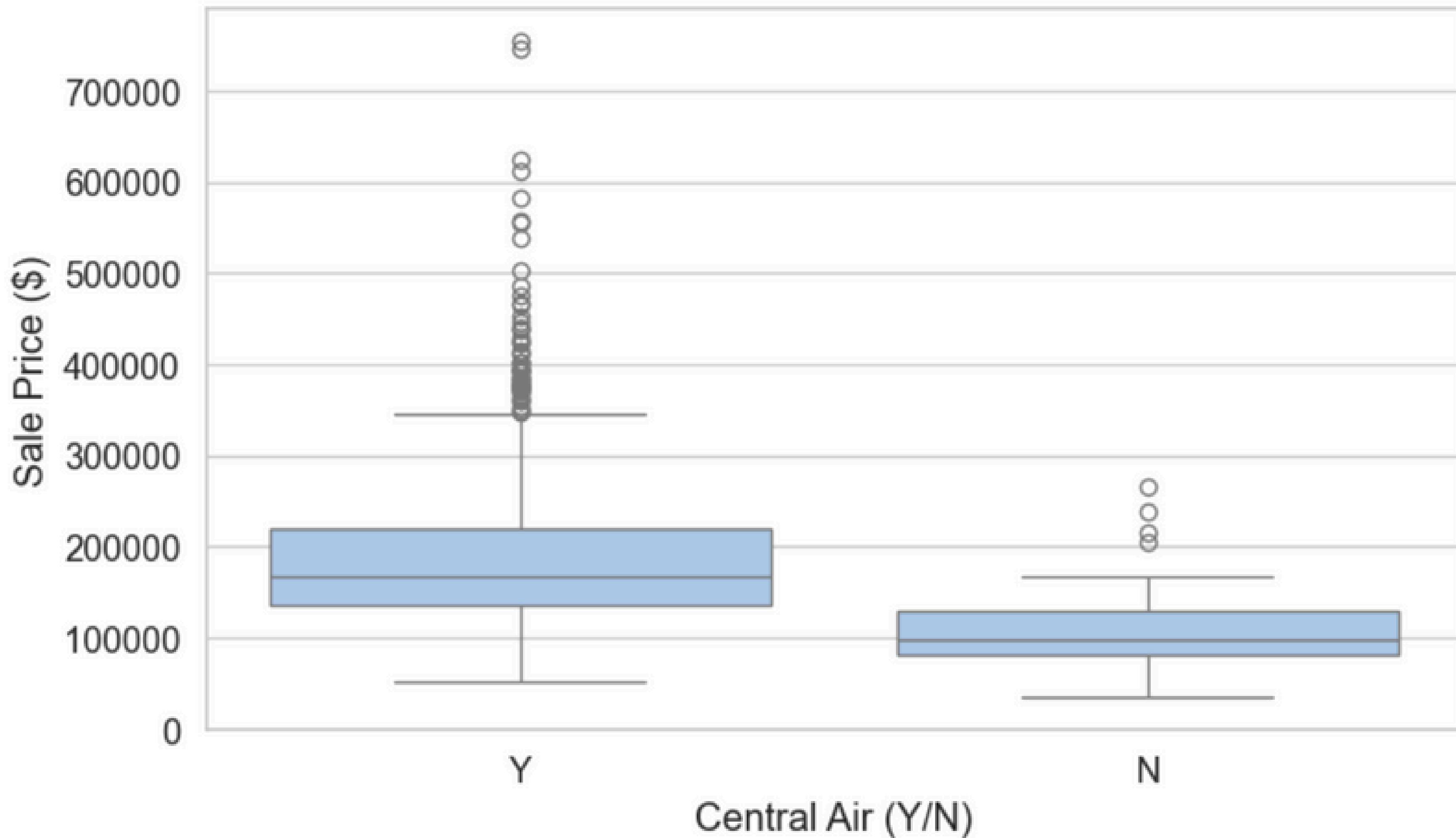
Price Distribution based on Swimming Pool Availability



### SalePrice vs. Garage Cars



## Sale Price vs Central Air



# Clustering Analysis

- Segmented the Market with K-Means

Used clustering to group homes based on size and price-related features.

- 3–4 Distinct Segments Identified

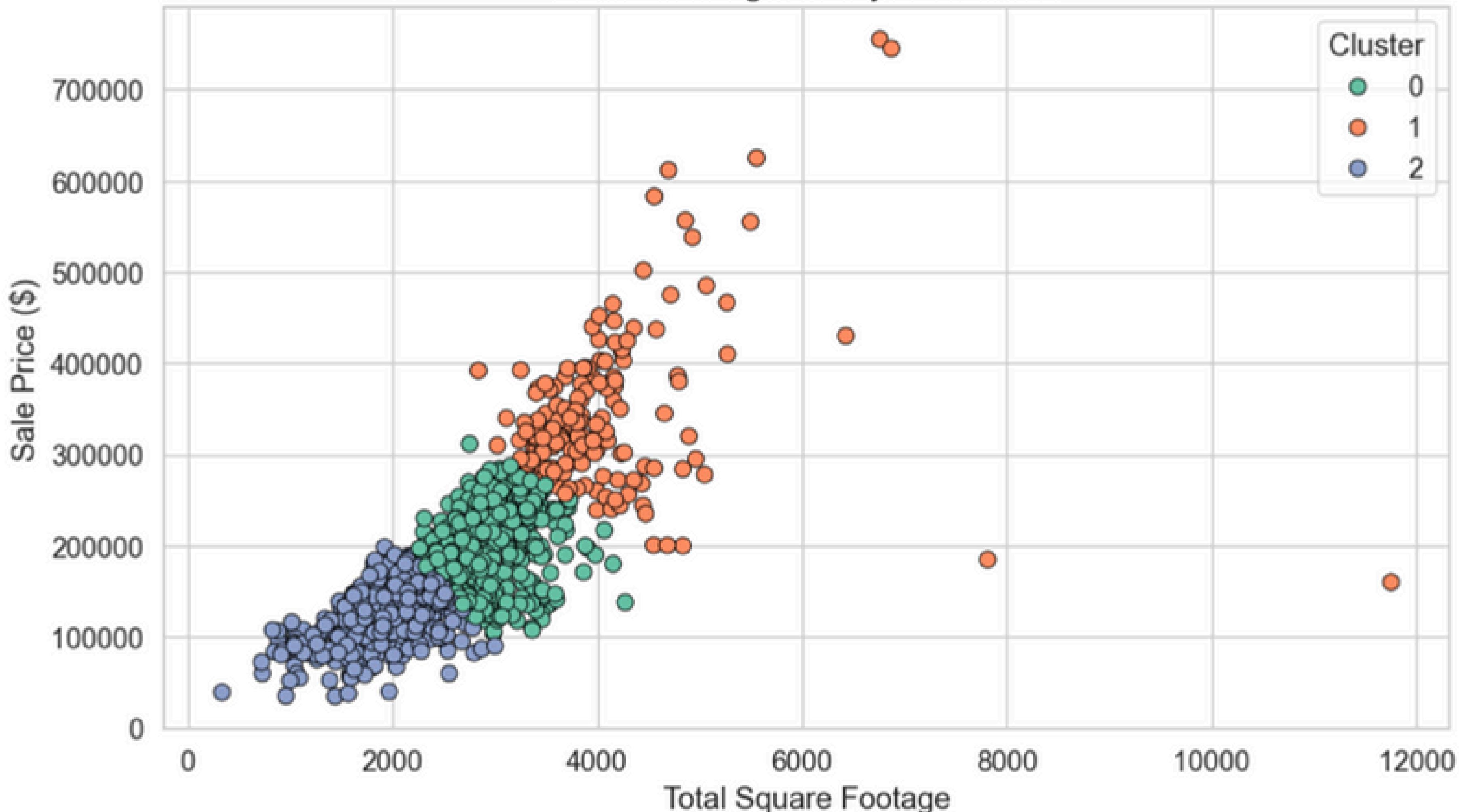
Clusters represent budget, mid-range, and premium property groups.

- Smarter Pricing Strategies

Helps tailor pricing, marketing, and investment strategies for each market segment.



## Market Segments by Size & Price





# Key Insights

## Size, Quality & Location Matter Most

These are the strongest predictors of a home's value.

## Smart Features = Better Insights

Calculated metrics like price\_per\_sqft help reveal pricing patterns.

## Trends & Amenities Add Value

Time of sale and features like garages, pools, and central air influence price.

## Clustering Unlocks Strategy

Grouping homes by size and price supports targeted pricing and marketing.

# Strategic Pricing & Investment Recommendations

## \$ Price Homes by Neighborhood

Use adjusted price per square foot tailored to each neighborhood's trends.

## 🏡 Invest in Quality Over Size

Focus on enhancing garage space and interior quality rather than just size.

## 🔍 Segment Homes with Clustering

Leverage clustering results to target specific buyer segments more effectively.

## JUL 17 Adapt to Seasonal Trends

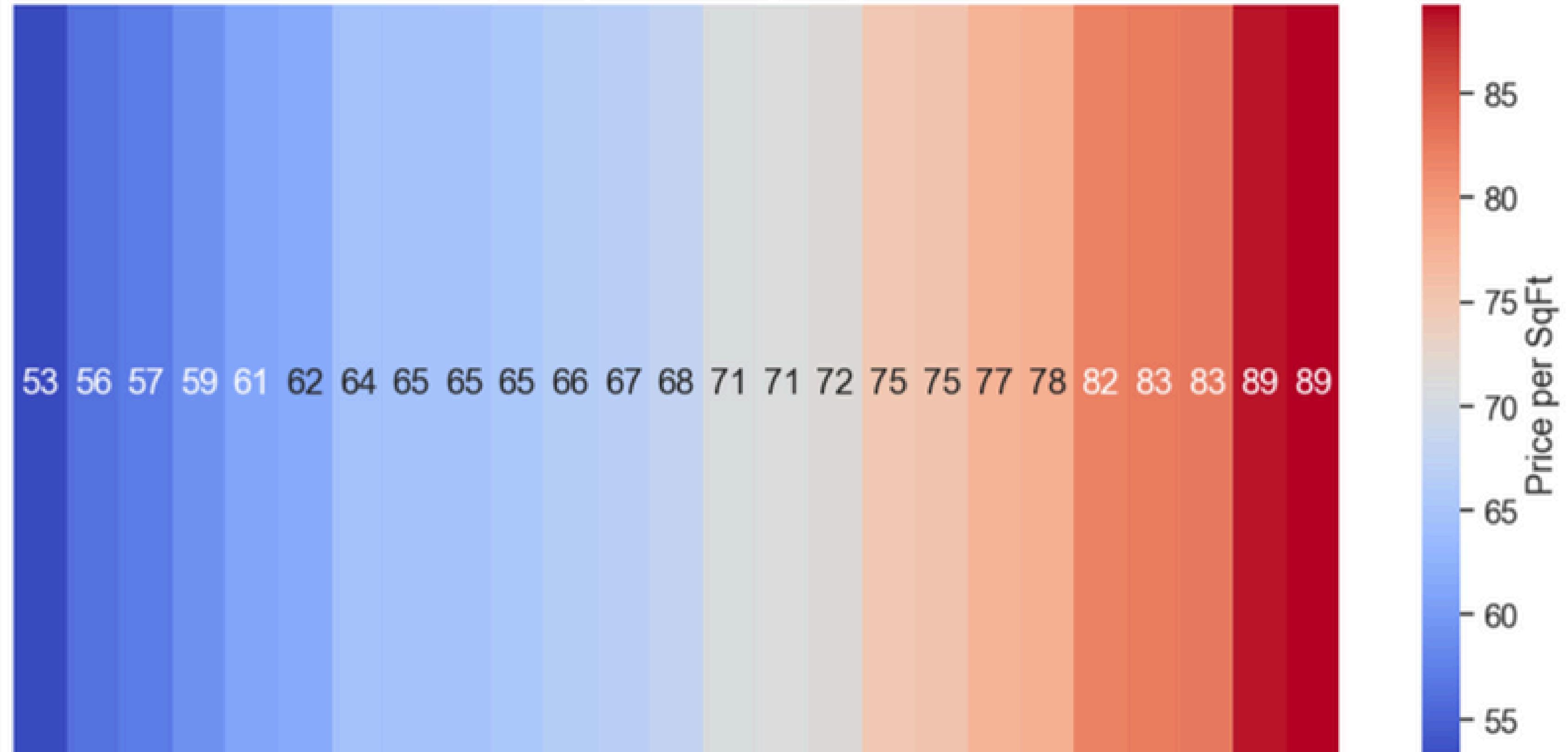
Monitor sales cycles and adjust prices based on seasonal demand.

## STRATEGIC PRICING & INVESTMENT RECOMMENDATIONS

Potential Investment & Recommendations for Real Estate

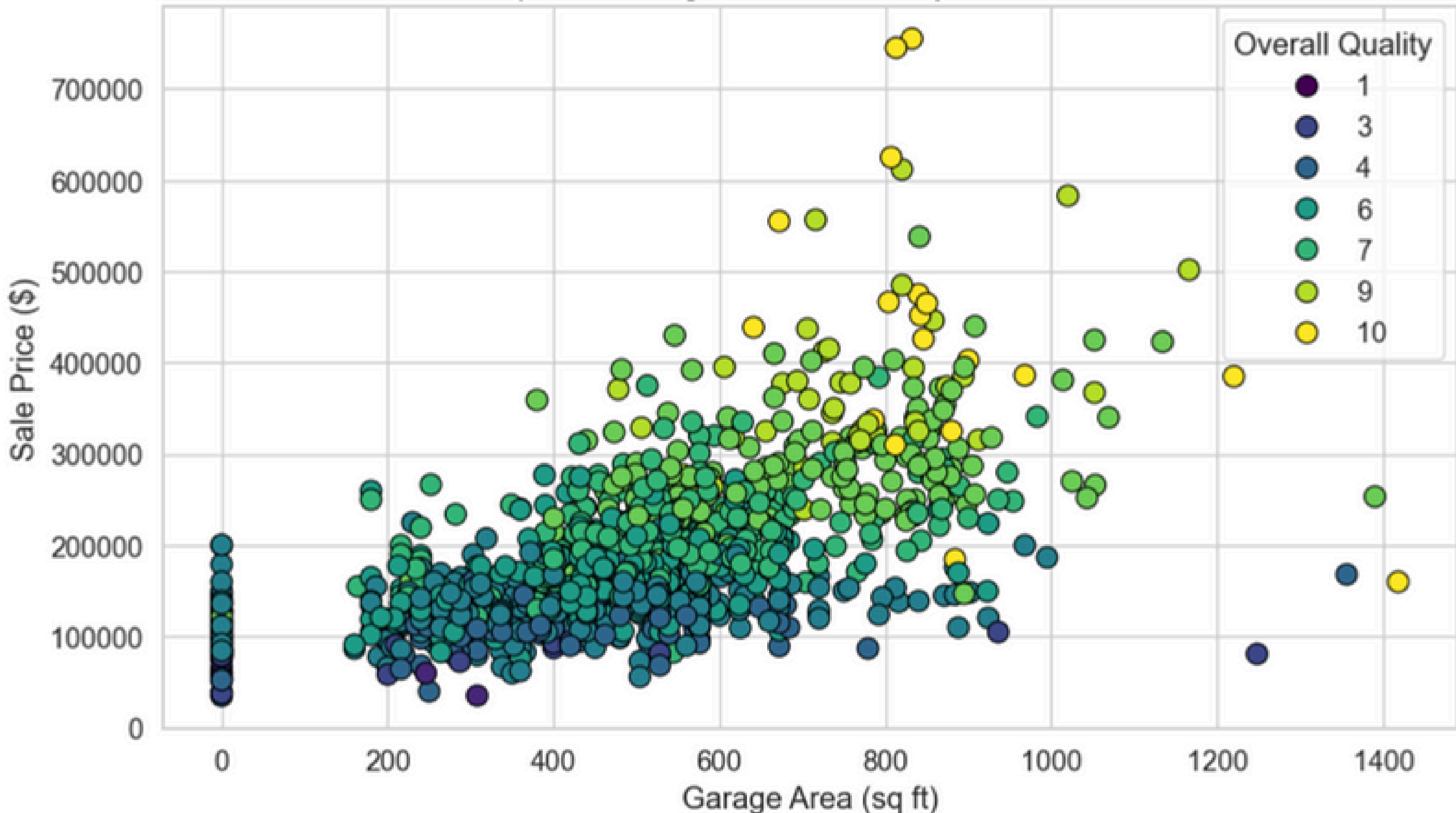


## □□ Price per Square Foot by Neighborhood

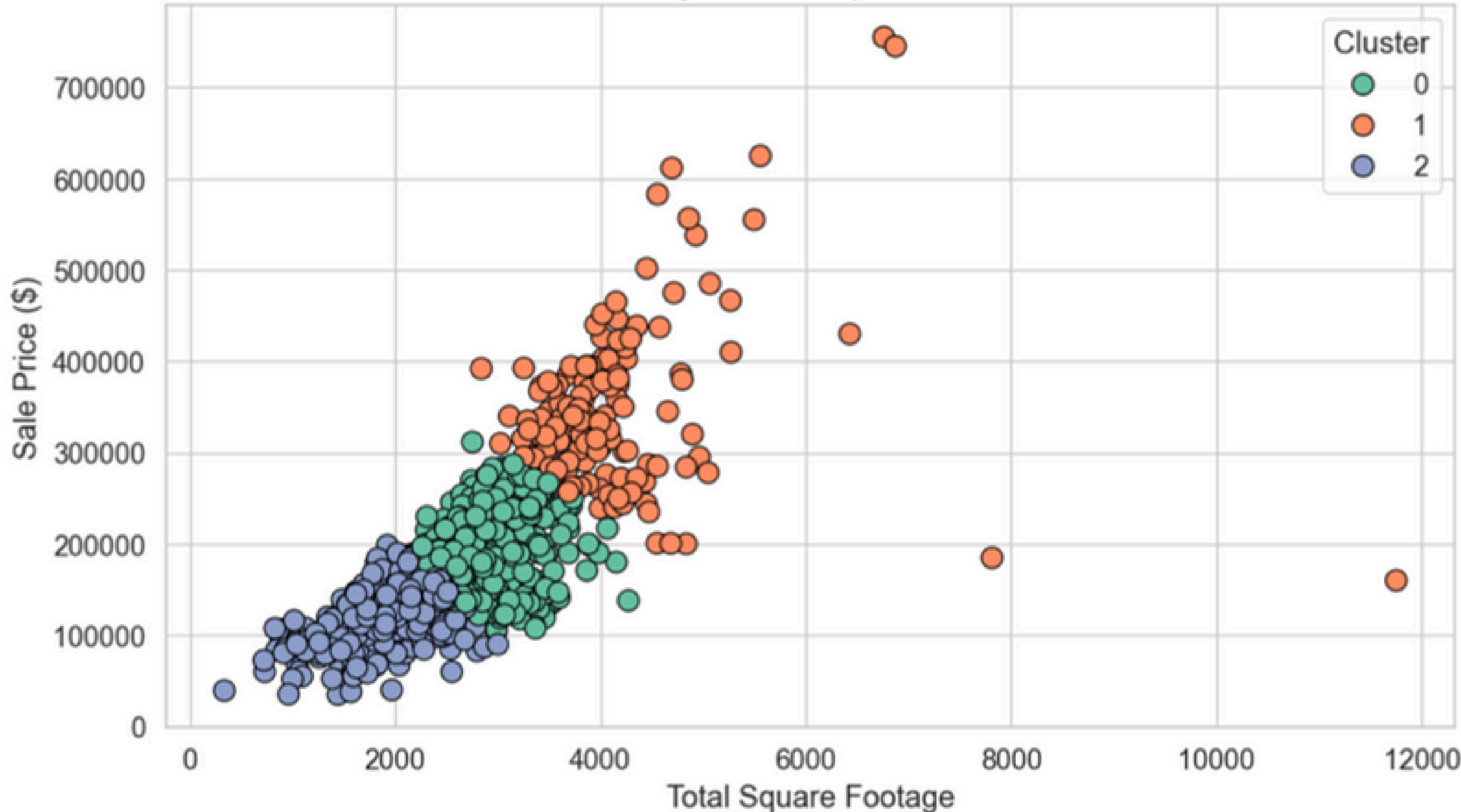


IDOTRR  
SWISU  
OldTown  
MeadowV  
BrDale  
Edwards  
Sawyer  
BrkSide  
NAmes  
NPkVll  
NWAmes  
Mitchel  
Blueste  
Blmgtn  
SawyerW  
ClearCr  
Crawfor  
CollgCr  
Gilbert  
Timber  
Veenker  
Somerset  
NoRidge  
NridgHt  
StoneBr

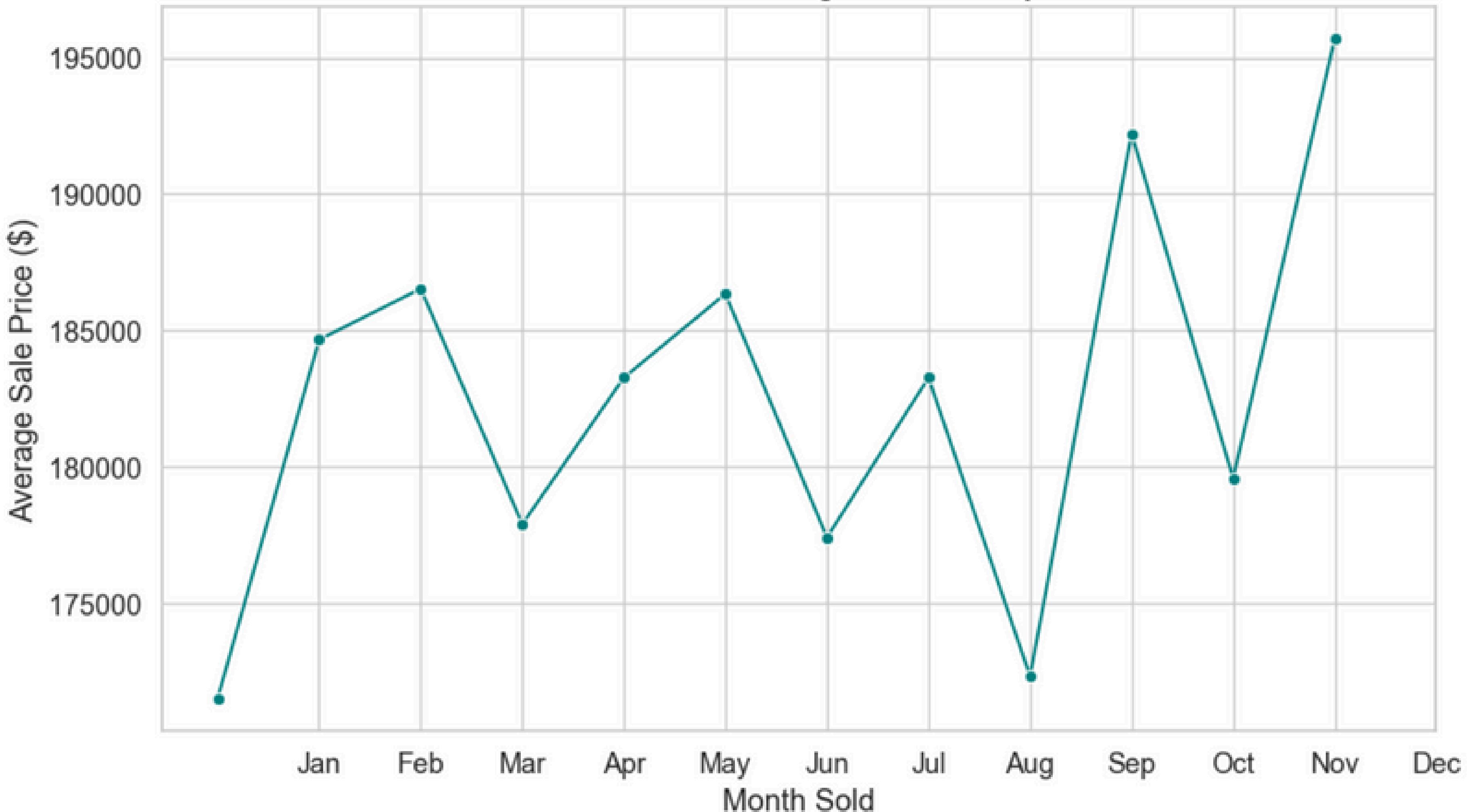
# Impact of Garage Area and Quality on Sale Price



# Home Segmentation by Size and Price



## Seasonal Trends: Average Sale Price by Month





# Thank you