

Exploratory Data Analysis (EDA) for Real Estate Pricing

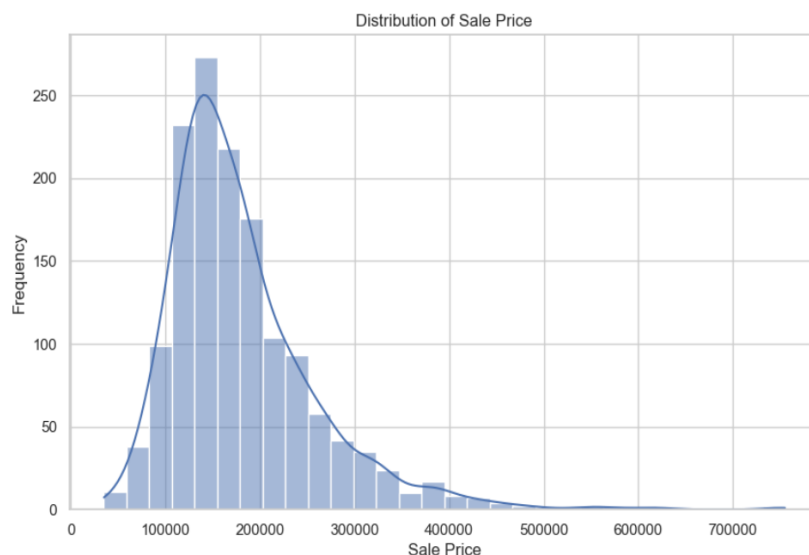
The real estate market is influenced by multiple factors including location, property size, amenities, and age. This project aims to perform an exploratory data analysis (EDA) on a housing dataset to uncover patterns, relationships, and insights that impact house pricing.

Dataset Description

The dataset contains information about houses, including numeric features like living area size, garage space, and sale price, as well as categorical features such as property type and location. Key features include SalePrice, GrLivArea, GarageArea, GarageCars, PoolArea, YearBuilt, YrSold, and MoSold.

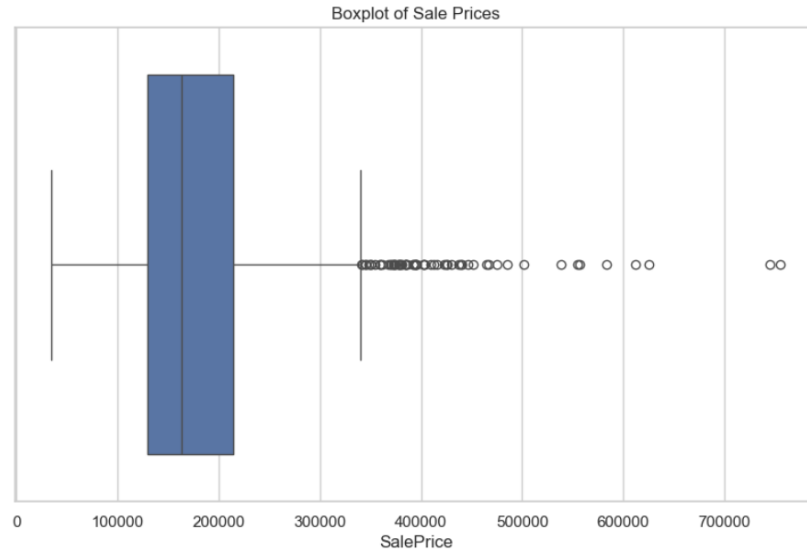
Findings and Insights

1. Distribution of Sale Price



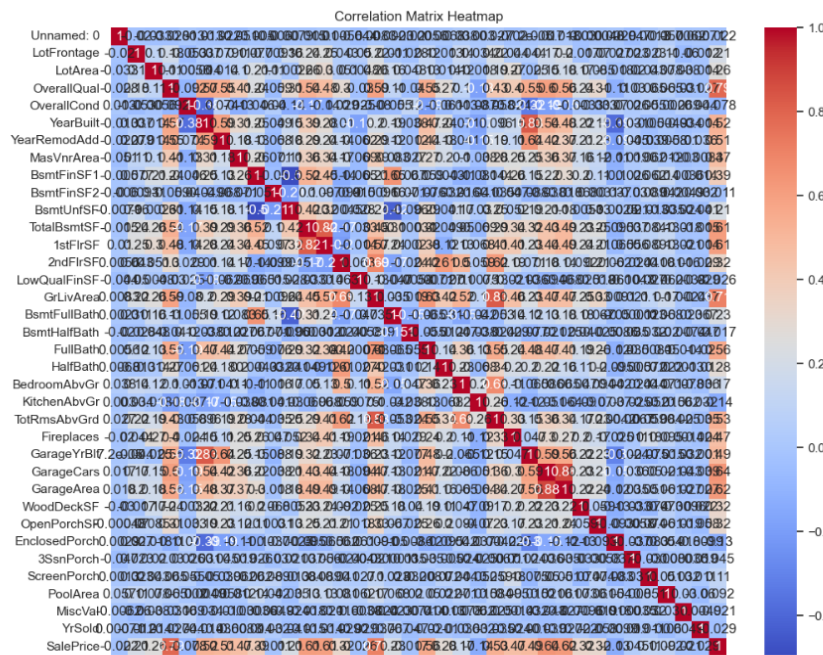
House prices are right-skewed, with most homes priced between 120,000 and 200,000. A few properties are extreme outliers with much higher prices.

2. Outlier Detection in Prices



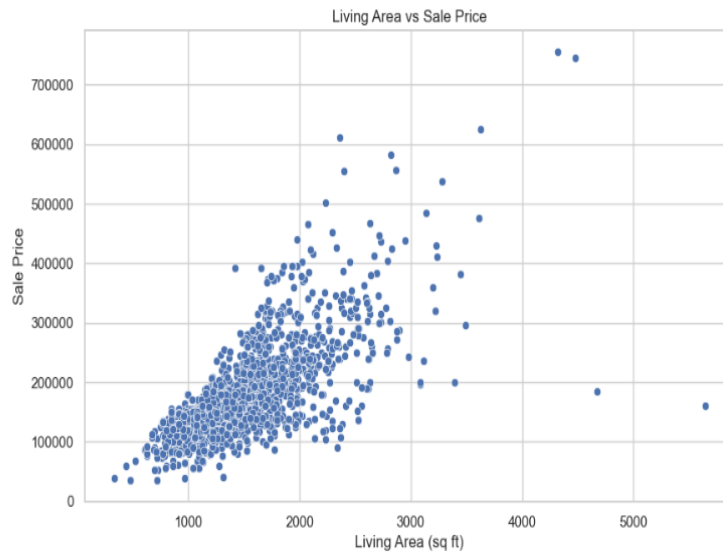
The boxplot confirms the presence of high-value outliers. Most houses are concentrated in the 120,000–250,000 range.

3. Feature Correlation



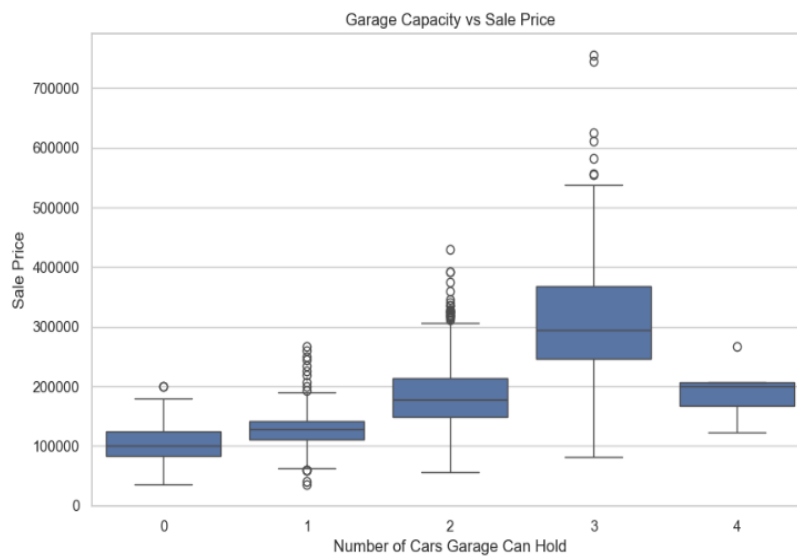
The heatmap shows strong positive correlations between SalePrice and features such as OverallQual, GrLivArea, and GarageCars.

4. Effect of Living Area



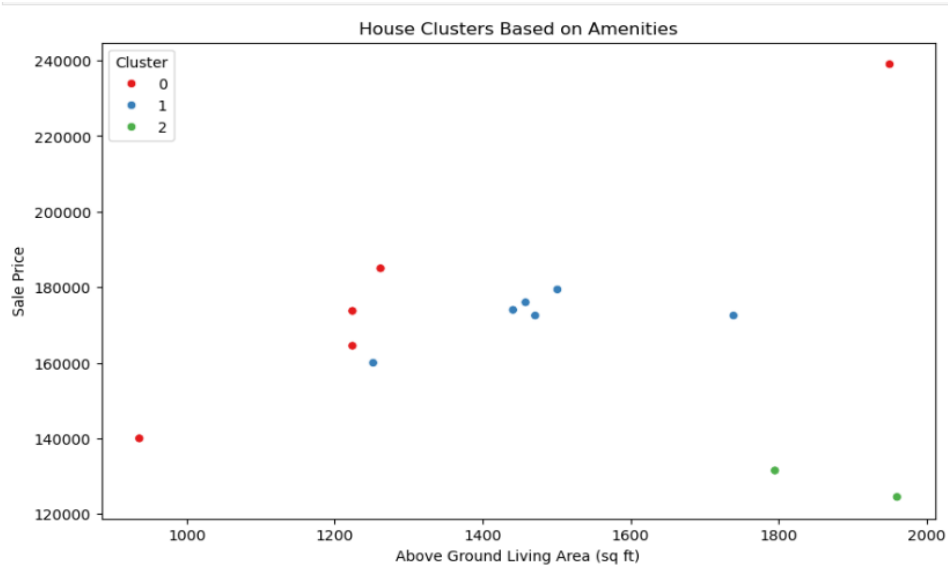
The scatter plot shows a clear positive relationship between living area and sale price. Larger homes tend to sell for higher prices.

5. Effect of Garage Capacity



Houses with larger garages generally command higher prices. Homes with 2–3 car garages have a noticeably higher median sale price.

6. Clustering of Houses



Clustering identified three groups: premium homes with high prices and superior amenities, average homes with mid-range pricing, and larger low-priced homes.

Conclusion

The analysis revealed that property size, quality, and amenities are the strongest factors influencing real estate prices. Clustering showed distinct property segments that can be targeted for better pricing strategies. These insights can guide developers, investors, and buyers in making data-driven decisions in the housing market.