

Exploratory Data Analysis (EDA)

1. Dataset Overview

The dataset contains 1460 rows and 81 columns. It includes a mix of numerical and categorical variables.

2. Missing Values Analysis

Several columns have missing values, with features like Alley, PoolQC, Fence, and MiscFeature having a high percentage of missing values.

LotFrontage, MasVnrType, and Garage-related features also contain missing values

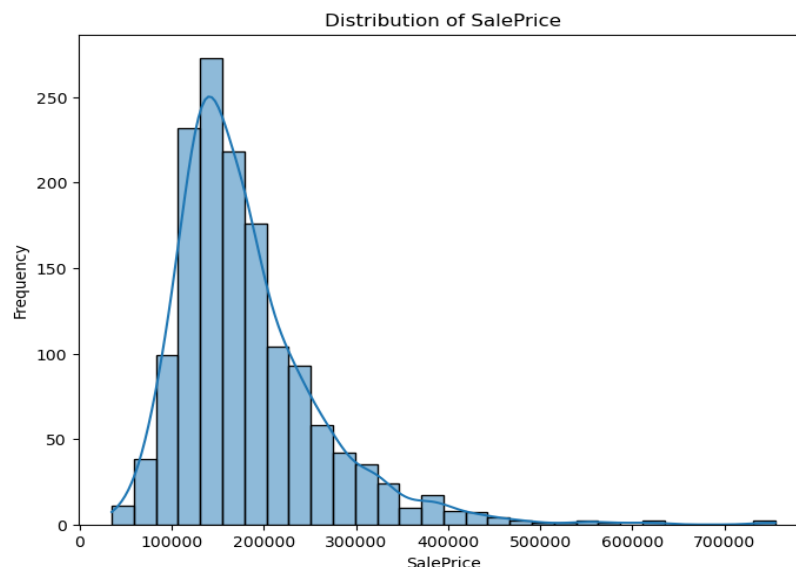
3. Summary Statistics

The average SalePrice is \$180,921 with a standard deviation of \$79,442. The minimum sale price is \$34,900, while the maximum is \$755,000. The distribution of SalePrice is right-skewed, meaning most houses are priced lower, with a few high-value outliers.

4. Correlation Analysis

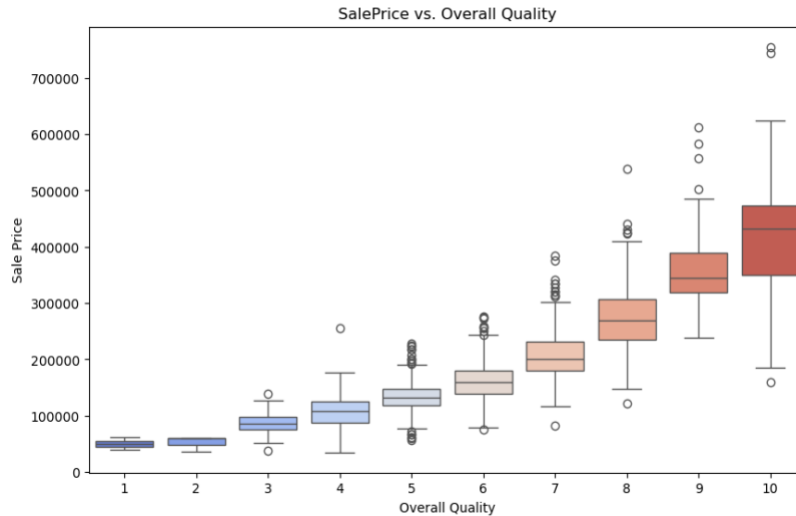
OverallQual, GrLivArea, and TotalBsmtSF strongly correlate with SalePrice, indicating that house quality, above-ground living area, and basement size significantly impact price. Features like GarageCars and 1stFlrSF also show positive correlations with price.

5. Key Visualizations

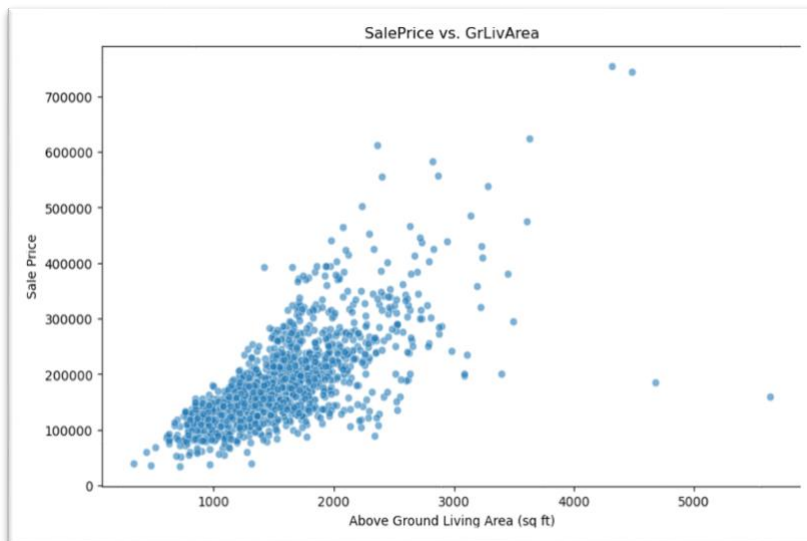


Sale Price Distribution: The histogram shows a skewed distribution, indicating potential transformation needs for modeling.

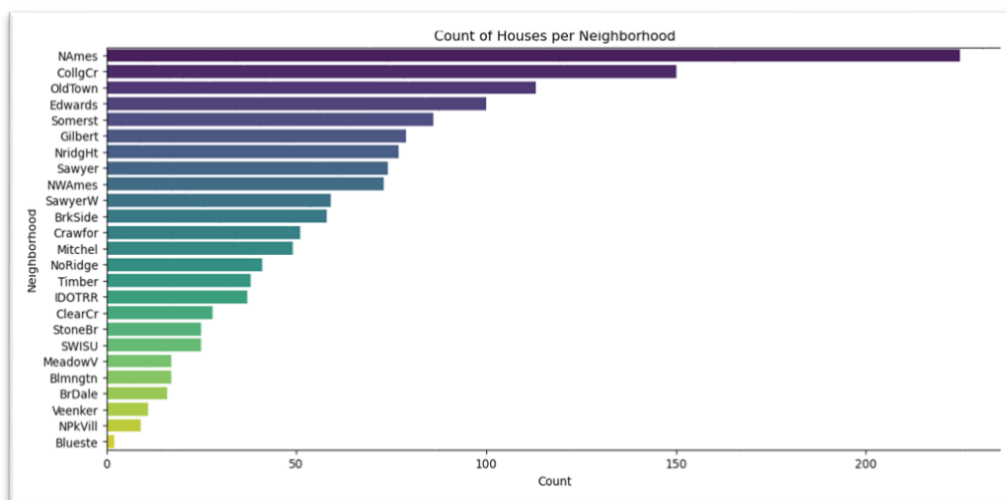
have



SalePrice vs. Overall Quality: Houses with higher quality ratings significantly higher prices.



Price vs. GrLivArea: There is a strong positive correlation between house size and price, with some outliers.



Neighborhood Distribution: Some neighborhoods have significantly more homes in the dataset, impacting location-based analysis.

