

# INSIGHTS&INFERENCES

## Descriptive Analysis:

### 1. Dataset Dimensions and Structure

- The dataset has **1460 rows** and **81 columns**.
- The columns contain various types of data including integers, floats, and categorical features (objects).

### 2. Missing Values Summary

- A total of **19 columns** have missing values.
- The columns with the highest percentages of missing values are:
  - **PoolQC** (99.52% missing)
  - **MiscFeature** (96.30% missing)
  - **Alley** (93.77% missing)
  - **Fence** (80.75% missing)
  - **FireplaceQu** (47.26% missing)
- Columns with missing values but less than 20% include **LotFrontage** (17.74%), which could be significant for analysis.

### 3. Summary Statistics for Numerical Columns

- Key statistics for some important variables:
  - **LotArea** (land area in square feet):
    - Mean: **10516.83**
    - Min: **1300**, Max: **215245**
    - 75th percentile: **11517**, indicating many properties have a relatively compact size.
  - **SalePrice** (target variable):
    - Mean: **180,921.20**
    - Median: **163,000**
    - Standard deviation: **79,442.50**
    - Min: **34,900**, Max: **755,000**

### 4. Distribution of the Target Variable (SalePrice)

- **Skewness:** The target variable has a positive skew (skewness = **1.88**), suggesting that most houses are priced below the mean, with a few very high-priced properties pulling the mean upwards.
- **Kurtosis:** The kurtosis value is **6.53**, indicating the presence of extreme values.

### 5. Correlation Analysis (Top 5 Correlations with SalePrice)

- Features most correlated with **SalePrice**:
  - **OverallQual** (0.79): A measure of overall material and finish quality.
  - **GrLivArea** (0.71): Above-grade (ground) living area.
  - **GarageCars** (0.64): Number of cars that fit into the garage.
  - **GarageArea** (0.62): Area of the garage in square feet.
  - **TotalBsmtSF** (0.61): Total square footage of the basement.
- These correlations suggest that quality, living space, and garage attributes are strong predictors of house prices.

## 6. Categorical Variable Overview

- The dataset contains a mix of categorical features with varying levels. Examples include:
  - **MSZoning**: Five categories (e.g., RL, RM, etc.)
  - **SaleCondition**: Six categories (e.g., Normal, Abnorml, etc.)
  - **Neighborhood**: 25 unique categories representing different neighborhoods.
- Understanding these distributions can help in segmenting and categorizing houses based on zones, neighborhoods, and conditions.

## 7. LotFrontage Analysis (Linear Feet of Street Connected to Property)

- LotFrontage has **259 missing values** (17.74%).
- Mean LotFrontage is **70.05 feet**, but values range significantly from **21 to 313 feet**.
- Given the importance of land frontage in real estate pricing, imputing these missing values using strategies like median values per neighborhood could be beneficial.

## 8. Year Built and Renovation Analysis

- **YearBuilt**:
  - Mean: **1971**, with most properties built between **1950 and 2000**.
  - Minimum year built: **1872**, indicating some very old properties.
- **YearRemodAdd**:
  - Indicates the year of last remodeling. Mean is **1984**, suggesting most properties have seen updates.
  - A comparison of **YearBuilt** and **YearRemodAdd** shows that many houses were remodeled years after the initial construction, which could influence pricing.

## 9. GrLivArea (Above Grade Living Area) Summary

- Mean: **1515.46 sq. ft.**, Median: **1442 sq. ft.**, showing that the living area is slightly right-skewed.
- Min: **334 sq. ft.**, Max: **5642 sq. ft.**, with the larger areas likely corresponding to high-value properties.
- This variable has a strong correlation (0.71) with **SalePrice**, confirming its predictive power.

## 10. Garage Features Analysis

- **GarageArea** and **GarageCars** both correlate well with **SalePrice** (0.62 and 0.64 respectively).

- **GarageYrBlt** shows a high number of missing values (81 entries), likely due to houses without a garage.
- On average, garages have space for **1.77 cars**, with the majority (84%) having room for either 1 or 2 cars.

## 11. Basement Features Overview

- The basement areas show a wide range:
  - **TotalBsmtSF**: Mean = **1057.43 sq. ft.**, Max = **6110 sq. ft.**
  - **BsmtFinSF1** (primary finished area): Mean = **443.64 sq. ft.**, Max = **5644 sq. ft.**
  - Many properties have no finished basement space, with 50% of houses having **0** sq. ft. for **BsmtFinSF2** (secondary finished area).

## 12. First and Second Floor Area Analysis

- **1stFlrSF** (first-floor area) has a mean of **1162.63 sq. ft.**, and the second-floor area (**2ndFlrSF**) has a mean of **347 sq. ft.**, indicating that many properties are single-story or have minimal second-floor space.
- 50% of properties have no second floor (median **2ndFlrSF** = 0).

## 13. Masonry Veneer Area and Type Analysis

- **MasVnrArea** (masonry veneer area):
  - Mean = **103.69 sq. ft.**, but the median is **0 sq. ft.**, indicating many properties do not have masonry veneer.
- **MasVnrType** shows:
  - **None**: 864 entries
  - **BrkFace** (brick face) veneer is most common among those with masonry.
  - Only **8** missing entries in this column.

## 14. Kitchen Quality and SalePrice Relationship

- Average **SalePrice** based on kitchen quality:
  - **Ex** (Excellent): **\$328,554**
  - **Gd** (Good): **\$212,116**
  - **TA** (Typical/Average): **\$139,963**
  - The difference shows a significant impact of kitchen quality on pricing.

## 15. Fireplaces Count and Quality Analysis

- **Fireplaces**:
  - Mean = **0.61** fireplaces per house; 50% of properties have at least one.
- **FireplaceQu** (quality):
  - 690 entries are missing (indicating no fireplace).
  - The most frequent quality for existing fireplaces is **Good** (380 entries).

## 16. Pool Area and Pool Quality Overview

- **PoolArea** has a mean of **2.76 sq. ft.**, indicating that most properties do not have pools (confirmed by the median of **0**).
- Only a few properties have pools, with **PoolQC** showing 1,453 missing values (99.5% of entries have no pool).

## 17. Paved Driveway (PavedDrive) and Its Impact

- The presence of a paved driveway correlates positively with price:
  - **Y (Yes):** Average price = **\$186,434**
  - **N (No):** Average price = **\$115,039**
- Indicates that a paved driveway is a significant positive feature.

## 18. Electrical Systems and Missing Values Impact

- Most properties use **SBrkr** (standard breaker), with **1334 entries**.
- Only **1** missing value in the **Electrical** column.
- Other types like **FuseA** (94) and **FuseF** (27) are less common.

## 19. Property Age Distribution and Its Effect on SalePrice

- Property age shows a negative correlation with **SalePrice** (**-0.523**), indicating that newer properties generally command higher prices.
- Properties tend to decrease in value as they age unless they undergo significant renovations.

## 20. Neighborhood-wise Analysis of House Prices

- Mean house prices vary significantly by neighborhood:
  - Lowest: **MeadowV** (\$98,576) and **IDOTRR** (\$100,124)
  - Highest: **NridgHt** (\$316,271) and **NoRidge** (\$335,295)
- Neighborhood plays a substantial role in property valuation, with certain areas consistently commanding premium prices.

## 21. Lot Shape and Impact on SalePrice

- Houses with irregular lot shapes tend to have higher prices:
  - **IR2** (Moderately irregular): **\$239,833**
  - **IR3** (Very irregular): **\$216,037**
  - **Reg** (Regular): **\$164,755**
- Irregular lots might offer unique advantages or larger spaces that are valued higher.

## 22. Street Type (Paved vs. Gravel) Analysis

- Properties on **paved streets** have a higher average price (**\$181,131**) compared to those on **gravel streets** (**\$130,191**).
- Indicates the importance of street quality in determining property value.

## 23. House Style and Its Influence on Pricing

- **2Story** houses have a higher average price (**\$210,052**) compared to **1Story** houses (**\$175,985**).
- The highest average price is seen in **2.5Fin** houses at **\$220,000**, possibly due to the additional finished half-story.

## 24. Building Type and SalePrice Relationship

- **1Fam** (single-family homes) have the highest average price (**\$185,764**).
- **Duplex** and **Twnhs** (townhouses) tend to have lower prices around **\$133,541** and **\$135,912**, respectively.

## 25. Foundation Type Analysis and Correlation with SalePrice

- **PConc** (Poured concrete) foundations correlate with higher average prices (**\$225,230**).
- Other foundation types like **Slab** have lower prices (**\$107,366**), indicating the preference for certain construction methods.

## 26. Exterior Condition and Its Impact on Pricing

- Houses with **Excellent (Ex)** exterior condition have a higher average price (**\$201,333**).
- **Fair (Fa)** and **Poor (Po)** conditions correlate with significantly lower prices (**\$102,595** and **\$76,500**), showing the influence of exterior maintenance on value.

## 27. Heating System Types and Their Effects

- Properties with **GasA** (Gas heating) systems have the highest average price (**\$182,021**), the most common and efficient type.
- Less efficient systems like **Wall** heating have much lower average prices (**\$92,100**).

## 28. Central Air Conditioning Presence and Its Influence

- Houses with **Central Air (Y)** have a significantly higher average price (**\$186,187**) compared to those without (**\$105,264**).
- This indicates that central air is a highly valued feature in the market.

## 29. Roof Style and SalePrice Correlation

- **Shed** and **Hip** roof styles show higher average prices (**\$225,000** and **\$218,877**), likely due to aesthetic appeal or architectural value.
- More common styles like **Gable** have a lower average price (**\$171,484**).

## 30. Sale Type and Condition Analysis

- **SaleType:**
  - **New** homes have the highest average price (**\$274,945**), indicating a premium for new constructions.
  - **COD** (Court order) and **WD** (Warranty deed) sales tend to have lower prices.
- **SaleCondition:**

- **Partial** sales (newly constructed homes partially completed) have the highest prices (\$272,292).
- **Abnorml** and **AdjLand** sales (abnormal or land-only adjustments) have the lowest values, indicating less favorable conditions.

### 31. Overall Quality (OverallQual) Distribution and Impact on SalePrice

- **SalePrice** increases significantly with higher overall quality:
  - **1**: \$50,150
  - **10** (highest quality): \$438,588
- This strong positive relationship emphasizes the importance of quality in pricing.

### 32. Overall Condition (OverallCond) Distribution and Correlation

- Houses with an overall condition of **5** have the highest average price (\$203,147), showing a peak at mid-range condition levels.
- Extreme low conditions (e.g., **1**) have prices as low as \$61,000, indicating poor condition significantly lowers value.

### 33. Garage Types and Their Influence on Pricing

- **BuiltIn** garages have the highest average price (\$254,752), showing a preference for integrated garage spaces.
- **CarPort** and **Detchd** garages show significantly lower average prices, indicating detached or open garages are less valuable.

### 34. Garage Finish Level Analysis

- **Fin** (Finished) garages have the highest average sale price (\$240,053).
- **Unf** (Unfinished) garages have the lowest at \$142,156, demonstrating that garage finishing level affects property value.

### 35. Garage Year Built and SalePrice Correlation

- The correlation between **GarageYrBltn** and **SalePrice** is **0.486**, indicating that newer garages are generally associated with higher property values.

### 36. Fence Quality Impact on SalePrice

- Properties with **GdPrv** (Good privacy) fences have the highest average price (\$178,927).
- **MnWw** (Minimum wood/wire) fences have lower average prices (\$134,286), showing that better-quality fencing adds value.

### 37. Alley Type and Its Effect on Pricing

- **Paved** alleys have a higher average sale price (\$168,001) compared to **Gravel** (\$122,219).
- This indicates that paved access routes are considered more valuable.

### 38. Roof Material Analysis

- **WdShngl** (Wood shingle) roofs have the highest average sale price (\$390,250), possibly due to aesthetic appeal.
- More common materials like **CompShg** (Composition shingles) have an average of \$179,804, showing that material choice affects pricing.

### 39. Kitchen Above Ground Area (GrLivArea) and Its Distribution

- The **GrLivArea** (above ground living area) ranges from **334 sq. ft.** to **5642 sq. ft.**, with a mean of **1515.46 sq. ft.**
- This suggests a significant variance in living space, which correlates positively with property value.

### 40. Total Bathrooms (Full + Half) and SalePrice Impact

- Properties with **3.5** bathrooms have the highest average sale price (\$417,266).
- As the total bathroom count increases, so does the average sale price, demonstrating the impact of bathroom availability on valuation.

#### SalePrice Distribution:

- Skewness: **1.88** (right-skewed), indicating many properties are priced below the mean with few high-value outliers.
- Kurtosis: **6.54**, showing the presence of extreme values.

#### LotArea Analysis:

- Skewness: **12.21**, demonstrating significant right skew due to a few very large properties.
- High kurtosis (**203.24**) implies outliers with extremely large lot sizes.

#### Year Built and Year Sold Analysis:

- Positive correlation (**0.52**) between **YearBuilt** and **SalePrice**, confirming that newer houses generally have higher prices.

#### Neighborhood Price Variation:

- Detailed statistics (min, max, median, mean, standard deviation, and IQR) per neighborhood illustrate the significant variability in pricing based on location.

#### Sale Type Comparison:

- Provides summary statistics for each sale type, revealing variability in pricing across different types of transactions.

#### Basement Area (TotalBsmtSF) Analysis:

- Strong correlation (**0.61**) with **SalePrice**, confirming that basement area contributes significantly to property value.

- Right-skewed distribution, indicating that most properties have smaller basements.

#### **Garage Area Analysis:**

- Strong correlation (**0.62**) with **SalePrice**, showing the importance of garage space.
- Skewness reveals a distribution with a few very large garage areas.

#### **Correlation Analysis with SalePrice:**

- Features like **OverallQual**, **GrLivArea**, and **GarageCars** show the highest correlations with **SalePrice**.

#### **Multicollinearity Check (VIF):**

- Several variables exhibit high VIF values, indicating multicollinearity issues that need addressing before modeling.

#### **LotFrontage Impact and Imputation:**

- Imputed missing values for **LotFrontage** based on neighborhood medians. This reduces variability and aligns distributions closer to neighborhood norms.

### **41. Clustering Analysis (Market Segmentation): Results**

The clustering analysis segmented the market into three distinct clusters:

1. **Cluster 0** (Mid-Range Houses):
  - **Mean SalePrice:** \$188,789
  - **Median SalePrice:** \$185,000
  - **Count:** 586 properties
  - These houses are balanced in terms of size and quality, making them appealing to mid-market buyers.
2. **Cluster 1** (Starter Homes):
  - **Mean SalePrice:** \$126,705
  - **Median SalePrice:** \$129,000
  - **Count:** 639 properties
  - These houses typically have smaller living spaces, fewer bathrooms, and lower overall quality, targeting budget-conscious buyers.
3. **Cluster 2** (Luxury Houses):
  - **Mean SalePrice:** \$308,723
  - **Median SalePrice:** \$290,000
  - **Count:** 235 properties
  - These properties have larger living areas, higher quality ratings, and better amenities, catering to the luxury segment

### **42.Elasticity**

The regression model reveals the sensitivity of **SalePrice** to changes in key features:



### 1. **GrLivArea (Above Ground Living Area):**

- Coefficient: **67.59** ( $p < 0.001$ )
- For each additional square foot, the sale price increases by approximately \$67.60.
- This feature is highly significant, confirming that living area size is a critical determinant of house value.

### 2. **GarageArea:**

- Coefficient: **133.48** ( $p < 0.001$ )
- An additional square foot in the garage area increases the price by about \$133.48, showing that garage space has a higher marginal impact compared to living space.

### 3. **TotalBath:**

- Coefficient: **16,370** ( $p < 0.001$ )
- Each additional bathroom (full or equivalent half) increases the sale price by around \$16,370, making it one of the most valuable additions.

The model explains about **62%** of the variation in **SalePrice** ( $R\text{-squared} = 0.619$ ), indicating that these features are major price drivers. The high significance of these variables suggests they provide strong ROI opportunities.

## 43. Neighbourhood Impact

### 1. **Premium Neighborhoods:**

- **NoRidge:** Mean price of \$335,295, highest standard deviation (121,412), indicating a diverse range of high-value properties.
- **NridgHt:** Mean price of \$316,271, with a slightly lower variance, showing consistent premium values.
- **StoneBr:** Mean price of \$310,499; also a premium area with a significant price range.

### 2. **Mid-Range Neighborhoods:**

- **Somerst:** Mean price of \$225,380, with a moderate standard deviation, indicating stable pricing.
- **ClearCr** and **Crawfor:** Both have mean prices above \$200,000, showing they are desirable but more affordable than the top premium neighborhoods.

### 3. **Budget Neighborhoods:**

- **MeadowV** and **IDOTRR:** Lowest mean prices, under \$105,000, making them the most affordable.
- **BrDale** and **OldTown** also fall into the lower price range, indicating these areas may have older or smaller properties.

The data shows clear distinctions between premium, mid-range, and budget neighborhoods, demonstrating how location is a significant determinant of property value.

## 44. Renovation Impact

The analysis shows the impact of remodeling on sale prices:

### 1. **Non-Remodeled Properties:**

- **Mean SalePrice:** \$182,584
- **Median SalePrice:** \$170,000
- **Standard Deviation:** \$70,334 (lower variability, indicating consistency)

- **Count:** 764 properties
- 2. **Remodeled Properties:**
  - **Mean SalePrice:** \$179,096
  - **Median SalePrice:** \$155,000
  - **Standard Deviation:** \$88,383 (higher variability, showing a wider range of outcomes)
  - **Count:** 696 properties

Interestingly, remodeled properties have a lower median and mean sale price, with higher variability. This suggests that while some renovations significantly increase value, not all remodeling efforts yield high returns. The type and quality of renovation, as well as other property characteristics, likely influence this outcome.

## 45.Interaction of Living Area and Basement Size on Sale Price

This analysis shows the impact of the interaction between above-ground living area and basement size on sale prices:

1. **Small GrLivArea ( $\leq 1000$  sq. ft.):**
  - Sale prices range from around \$84,631 (small basement) to \$128,000 (large basement).
  - Limited data for very large or huge basements, indicating such configurations are uncommon in small houses.
2. **Medium GrLivArea (1000 - 2000 sq. ft.):**
  - Prices increase as basement size grows:
    - Small basement: \$145,594
    - Medium basement: \$157,029
    - Very large basement: \$253,652
  - This indicates that for medium-sized houses, increasing basement area has a substantial impact on price.
3. **Large GrLivArea (2000 - 3000 sq. ft.):**
  - A similar trend, with very large basements pushing prices over \$344,407.
  - Huge basements correlate with the highest values in this category, averaging \$363,744.
4. **Very Large GrLivArea (3000 - 4000 sq. ft.):**
  - Properties with very large or huge basements achieve prices exceeding \$400,000, showing that in larger homes, expansive basements significantly enhance value.
5. **Huge GrLivArea ( $> 4000$  sq. ft.):**
  - Limited data, but houses with huge basements reach up to \$461,188, indicating that these configurations command premium prices.

## 46.Garage Size and Sale Price Relationship Across Property Types

### Single-Family Homes (1Fam):

- Properties with no garage have an average price of **\$96,003**.
- Properties with **1-car** garages increase to **\$130,258**, while **2-car** garages average **\$187,354**.

- Homes with **3-car** garages have the highest average, **\$317,054**, showing a significant premium for larger garages.

#### **Two-Family Conversions (2fmCon):**

- Prices for properties without garages are around **\$121,072**.
- A slight increase for properties with **1-car** garages (**\$121,500**) and **2-car** garages (**\$134,750**).
- Properties with **4-car** garages show a much higher price (**\$168,000**), though these are rare.

#### **Duplex:**

- Prices range from **\$131,117** for no garages to **\$137,500** for **3-car** garages.
- Properties with **4-car** garages reach over **\$200,000**, showing that larger garage capacity in duplexes can significantly increase value.

#### **Townhouses (Twnhs and TwnhsE):**

- **Twnhs** properties with no garage have the lowest prices (**\$89,200**).
- Prices increase notably with **2-car** garages (**\$167,033**).
- **TwnhsE** properties with **2-car** garages show even higher prices, reaching an average of **\$189,195**.

### **47.Effect of Property Age on Sale price**

#### **0-10 Years Old:**

- **Mean SalePrice:** \$234,567
- Newer properties command higher prices, indicating a premium for newer constructions.
- These properties also exhibit higher variability, suggesting that newer homes can vary greatly in size and features.

#### **11-20 Years Old:**

- **Mean SalePrice:** \$234,833, similar to the 0-10 range.
- This indicates that homes up to 20 years old retain high value, assuming they are well-maintained or modernized.

#### **21-40 Years Old:**

- **Mean SalePrice:** \$158,938, showing a significant drop.
- As homes age beyond 20 years, prices start to decline, likely due to aging infrastructure and design styles.

#### **41-60 Years Old:**

- **Mean SalePrice:** \$144,104, continuing the downward trend.
- These properties might require updates or renovations to maintain value.

#### **61-100 Years Old:**

- **Mean SalePrice:** \$130,817, indicating further price reduction.
- Homes in this age range may have historical value but might need significant modernization.

#### **100+ Years Old:**

- **Mean SalePrice:** \$153,118, a slight uptick from the previous age group.
- This suggests that very old homes may have unique architectural or historical features that retain value.

## **48.Impact of House Style on Market Segment Value**

#### **1Story (Single-Story Homes):**

- **Mean SalePrice:** \$175,985
- This is the most common house style, showing moderate pricing with a large variation, indicating a wide range in quality and features.

#### **2Story (Two-Story Homes):**

- **Mean SalePrice:** \$210,052
- Two-story houses generally command higher prices, likely due to the increased living space they provide.

#### **1.5Fin (One and a Half Stories, Finished):**

- **Mean SalePrice:** \$143,117
- Lower average prices, indicating that this style might not be as valued as full two-story or single-story homes.

#### **2.5Fin (Two and a Half Stories, Finished):**

- **Mean SalePrice:** \$220,000
- Despite its rarity, this style shows high value, suggesting that when well-finished, these homes are desirable.

#### **SFoyer (Split Foyer) and SLvl (Split Level):**

- **SFoyer:** \$135,074
- **SLvl:** \$166,703
- These split-level designs have lower average prices, making them more affordable options.

## **49.Seasonal Effect: Influence of Sale Month on Pricing**

#### **Highest Average Prices:**

- **September:** \$195,683, with a median of \$185,900, indicating that early fall tends to have higher prices.
- **November:** \$192,211, showing that late fall may also see higher-value transactions.
- These months likely coincide with market conditions where supply and demand favor sellers, driving up prices.

### **Lowest Average Prices:**

- **April:** \$171,503, with the lowest median of \$147,000, suggesting that spring may be a more competitive period for buyers.
- **May** and **June** have similarly lower average prices compared to the fall months, indicating that increased supply might moderate prices.

### **Summer and Winter:**

- **July** and **August** have slightly higher prices (\$186,331 and \$184,652, respectively), reflecting stable market conditions.
- **December:** Although winter generally has lower activity, December shows higher average prices (\$186,519), suggesting that buyers looking for year-end deals may still pay competitive prices.

## **50. Analysis of House Prices Across Zoning Categories**

### **FV (Floating Village Residential):**

- **Mean SalePrice:** \$214,014, the highest among all zones.
- Indicates that properties in this zoning category are premium, likely due to desirable features or locations.

### **RL (Residential Low Density):**

- **Mean SalePrice:** \$191,005, with the largest count (1151 properties).
- This zone represents most properties, showing a wide range of prices and variability due to different property types within it.

### **RM (Residential Medium Density):**

- **Mean SalePrice:** \$126,317, indicating mid-range pricing.
- These properties are generally more affordable compared to RL or FV zones.

### **RH (Residential High Density):**

- **Mean SalePrice:** \$131,558, slightly higher than RM.
- This suggests that higher-density residential areas may still offer some premium value, likely due to urban proximity or amenities.

### **C (Commercial):**

- **Mean SalePrice:** \$74,528, the lowest of all, indicating that properties zoned for commercial use are less valuable as residential homes.

## **51. Impact of Exterior Quality and Condition on Sale Price**

### **Excellent Quality (Ex):**

- **ExterCond = TA (Typical/Average):** Mean price is **\$375,551**, with high variability due to 48 properties. This combination achieves the highest average value.
- Other combinations with **Ex** have smaller sample sizes but still show premium pricing (e.g., **Ex + Gd**: \$250,433).

#### **Good Quality (Gd):**

- **ExterCond = TA:** Mean price is **\$231,694** across 453 properties, indicating that good quality with average condition maintains high value.
- **ExterCond = Gd:** Slightly higher mean price of **\$230,848**, though with fewer properties (35).

#### **Typical Quality (TA):**

- **ExterCond = TA:** Most common combination with 773 properties and an average price of **\$145,018**.
- **ExterCond = Gd:** A slight increase in average price to **\$146,556**, showing that a better condition improves value even with typical exterior quality.

#### **Fair Quality (Fa):**

- Lowest prices are seen with **Fa** quality and condition combinations, such as **Fa + Fa** (\$63,717) and **Fa + TA** (\$106,187).

## **52. Influence of Lot Shape and Configuration on Pricing**

1. **IR2 (Moderately Irregular):**
  - **Mean SalePrice:** \$239,833, the highest among all shapes.
  - This suggests that moderately irregular lots may provide unique layouts or additional space that increase property value.
2. **IR3 (Very Irregular):**
  - **Mean SalePrice:** \$216,037, indicating that even very irregular lots can yield higher values.
  - The small sample size (10 properties) suggests that these configurations are less common but valuable.
3. **IR1 (Slightly Irregular):**
  - **Mean SalePrice:** \$206,102, showing a slight premium over regular lots.
  - With a larger count (484), it indicates that slight irregularity is fairly common and valued.
4. **Reg (Regular):**
  - **Mean SalePrice:** \$164,755, the lowest among the categories.
  - Although regular lots are the most common (925 properties), they typically command lower prices compared to irregular configurations.

## **53. Effect of Roof Style and Material on House Value**

#### **Gable Roof with Composition Shingles (CompShg):**

- **Mean SalePrice:** \$170,457, the most common combination (1134 properties).
- This configuration represents typical construction practices and yields mid-range prices.

### **Hip Roof with Composition Shingles:**

- **Mean SalePrice:** \$218,602, indicating that hip roofs provide a premium over gable roofs.
- The higher standard deviation suggests a broad price range, likely influenced by other features.

### **Gable Roof with Wood Shingles (WdShngl):**

- **Mean SalePrice:** \$412,700, showing a significant premium for this combination, though it is rare (5 properties).
- This suggests that unique roofing materials combined with a common roof style can yield high value.

### **Flat Roofs with Membran and Tar&Grv Materials:**

- **Mean SalePrices:** \$241,500 and \$187,447, respectively.
- These styles show variation but generally achieve moderate to high prices depending on materials.

### **Shed and Mansard Roofs with Wood Shake (WdShake):**

- **Mean SalePrices:** \$260,000 and \$216,000, indicating that less common roof styles with premium materials can achieve higher value

## **54.Interaction Between Kitchen Quality and Overall House Quality**

### **Excellent Kitchen Quality (Ex):**

- For **Overall Quality (OverallQual) = 10**: Mean price is **\$433,721**, the highest combination, indicating a premium for top-quality kitchens in the best overall houses.
- Prices drop slightly with lower overall quality (e.g., **OverallQual = 8**: \$305,271), showing the importance of both kitchen and overall house quality in determining value.

### **Good Kitchen Quality (Gd):**

- The most common configuration is **OverallQual = 7** with a mean price of **\$211,644** (244 properties).
- As overall quality increases, so does the price (e.g., **OverallQual = 10**: \$462,927), but it also highlights the variability (higher standard deviation).

### **Typical Kitchen Quality (TA):**

- For mid-range overall quality (**OverallQual = 6**), the mean price is **\$153,807**.
- The pattern is consistent: higher overall quality paired with average kitchen quality results in moderate pricing.

### **Fair Kitchen Quality (Fa):**

- For **OverallQual = 4 or lower**, prices remain low (e.g., **\$92,833**), emphasizing that both kitchen and overall quality must be improved together to significantly impact value.

## 55.Effect of Pool Area and Quality on House Pricing

### Excellent Pool Quality (Ex):

- **Large Pools:** Mean price is **\$490,000**, but only 2 properties fall into this category, indicating that while rare, such features can command very high prices.
- No properties in the dataset have small or medium pools with excellent quality.

### Fair Pool Quality (Fa):

- **Large Pools:** Mean price is **\$215,500**, with 2 properties.
- This indicates a significant drop compared to excellent pool quality, suggesting that even with a large pool, poor quality does not add much value.

### Good Pool Quality (Gd):

- **Large Pools:** Mean price is **\$201,990**, with 3 properties.
- Despite the good rating, the price is similar to the fair quality, implying that for pools, size might matter less if quality is not premium.

## 56.Relationship Between Heating System Types and Sale Price

### GasA (Gas Forced Air):

- **Mean SalePrice:** \$182,021, the most common system (1428 properties).
- This system offers mid-range pricing with significant variability, indicating that it's standard for many property types.

### GasW (Gas Hot Water):

- **Mean SalePrice:** \$166,632, with only 18 properties.
- This type of heating system has a slightly lower average value than GasA, suggesting it may be less efficient or desirable.

### Other Systems (Grav, OthW, Wall):

- Prices range from **\$75,271** (Gravity Heating) to **\$125,750** (Other Water-based).
- These systems are less common and typically associated with lower sale prices, possibly due to their outdated or less efficient nature.

## 57.Garage Finish Level Analysis

### Finished (Fin):

- **Mean SalePrice:** \$240,053, the highest among garage finish levels.
- This indicates that fully finished garages add significant value, especially in mid to high-range properties.

### Rough-Finished (RFn):



- **Mean SalePrice:** \$202,069, lower than fully finished garages.
- These garages still command a premium over unfinished ones, indicating that even partial finishes are valued.

#### **Unfinished (Unf):**

- **Mean SalePrice:** \$142,156, the lowest among categories.
- This suggests that properties with unfinished garages are typically valued lower, showing the importance of completing this space to increase property value.

## **58.Impact of Paved Driveway on Sale Price**

#### **Yes (Paved):**

- **Mean SalePrice:** \$186,434, indicating a significant premium for properties with paved driveways.
- This category has the largest count (1340 properties), suggesting that paved driveways are common and desirable.

#### **Partial (P):**

- **Mean SalePrice:** \$132,330, indicating a lower price compared to fully paved driveways.
- These properties are less common (30 properties) and show moderate value.

#### **No (Unpaved):**

- **Mean SalePrice:** \$115,039, the lowest among the categories.
- This suggests that unpaved driveways are associated with lower property values, as they are likely perceived as less convenient.

## **59.Interaction of Fireplace Count and Quality on Property Value**

#### **Single Fireplace:**

- **Excellent Quality (Ex):** Highest mean price of \$311,313 (19 properties), demonstrating that even a single high-quality fireplace can significantly add value.
- **Good Quality (Gd):** Mean price of \$223,404 (324 properties), showing that good-quality fireplaces are valuable but not as impactful as excellent ones.

#### **Two Fireplaces:**

- **Excellent Quality:** Mean price of \$457,539 (4 properties), the highest observed combination, indicating a premium for homes with multiple high-quality fireplaces.
- **Good Quality:** Mean price of \$245,937 (54 properties), also showing a premium, though not as high as for excellent quality.

#### **Three Fireplaces:**

- **Various Qualities:** Limited data, but mean price varies based on quality (e.g., \$360,000 for excellent, \$345,000 for typical).

## 60.Effect of Building Type on Sale Price

### Single-Family Detached (1Fam):

- **Mean SalePrice:** \$185,764, the highest and most common category (1,220 properties).
- This type commands premium pricing and is the most represented in the dataset, reflecting its popularity.

### Townhouse End Unit (TwnhsE):

- **Mean SalePrice:** \$181,959, close to single-family homes, suggesting that end-unit townhouses offer comparable value due to privacy and size benefits.

### Two-Family Conversions (2fmCon):

- **Mean SalePrice:** \$128,432, indicating mid-range pricing with moderate variability.
- This type is typically more affordable, catering to a different market segment.

### Townhouses Inside Unit (Twnhs):

- **Mean SalePrice:** \$135,912, slightly higher than duplexes but lower than end units, showing that privacy and layout differences affect value.

### Duplex:

- **Mean SalePrice:** \$133,541, the lowest among building types.
- Duplexes are generally more budget-friendly and have less demand compared to single-family homes.

## 61.Influence of Foundation Type on Property Value

### Poured Concrete (PConc):

- **Mean SalePrice:** \$225,230, the highest among foundation types, indicating a premium for properties with this modern and durable foundation.
- This type has the largest count (647 properties), showing its popularity.

### Stone Foundation:

- **Mean SalePrice:** \$165,959, indicating that stone foundations, although less common, can add value, likely due to their aesthetic or historical appeal.

### Cinder Block (CBlock):

- **Mean SalePrice:** \$149,806, showing mid-range pricing.
- This foundation type is common (634 properties) and provides a stable, albeit less premium option compared to poured concrete.

### Brick and Tile (BrkTil):

- **Mean SalePrice:** \$132,291, indicating lower value compared to concrete or stone foundations.

#### **Slab:**

- **Mean SalePrice:** \$107,366, the lowest, reflecting its association with smaller or budget properties.

## **62. Analysis of Sale Condition on House Prices**

#### **Partial (Partial Sale):**

- **Mean SalePrice:** \$272,292, the highest among sale conditions.
- This category often indicates new or recently completed construction, which commands a premium due to the property's newness and features.

#### **Normal (Typical Sale):**

- **Mean SalePrice:** \$175,202, the most common condition (1,198 properties).
- This represents standard market conditions, yielding mid-range pricing.

#### **Abnormal Sale (Abnormal):**

- **Mean SalePrice:** \$146,527, indicating a discount associated with properties sold under special circumstances (e.g., foreclosure, short sale).

#### **Family Sale (Family):**

- **Mean SalePrice:** \$149,600, typically reflecting sales between family members, which might result in below-market pricing.

#### **AdjLand (Adjoining Land Purchase):**

- **Mean SalePrice:** \$104,125, the lowest, showing that these sales focus on land rather than property value.

## **63. Relationship Between Lot Configuration (Corner vs. Inside) and Pricing**

#### **Cul-de-Sac (CulDSac):**

- **Mean SalePrice:** \$223,855, the highest among configurations, likely due to privacy and lower traffic, making these lots desirable.
- These properties also have high variability, indicating a range of house sizes and features.

#### **Inside Lots (Inside):**

- **Mean SalePrice:** \$176,938, the most common configuration (1,052 properties).
- Inside lots have a lower average price compared to corner and cul-de-sac lots, possibly due to less advantageous positioning.

#### **Corner Lots:**

- **Mean SalePrice:** \$181,623, slightly higher than inside lots.
- Corner lots often provide more space and exposure, which can add value.

#### **FR3 (Frontage on Three Sides):**

- **Mean SalePrice:** \$208,475, though very rare (4 properties).
- These configurations yield higher prices, likely due to additional exposure and space.

### **64.Effect of Presence of Central Air Conditioning on Sale Price**

#### **With Central Air (Y):**

- **Mean SalePrice:** \$186,187, substantially higher than properties without it.
- This suggests that central air is a highly valued feature, as it is associated with higher property values.

#### **Without Central Air (N):**

- **Mean SalePrice:** \$105,264, the lowest pricing segment.
- This indicates that properties lacking central air typically sell for much less, likely due to decreased comfort or lower quality of amenities.

### **65.Impact of Fence Quality on Property Value**

#### **Good Privacy Fence (GdPrv):**

- **Mean SalePrice:** \$178,927, the highest among fence types, indicating that privacy fences contribute significantly to property value.
- This type provides the most value, likely due to the enhanced privacy and security it offers.

#### **Good Wood Fence (GdWo):**

- **Mean SalePrice:** \$140,379, lower than good privacy fences.
- While adding value, wood fences without full privacy have a lesser impact compared to fully private fences.

#### **Minimum Privacy Fence (MnPrv):**

- **Mean SalePrice:** \$148,751, higher than minimum wood fences but less than good quality fences.
- This indicates that even a basic level of privacy in fencing can positively influence pricing.

#### **Minimum Wood Fence (MnWw):**

- **Mean SalePrice:** \$134,286, the lowest.
- Such fences are likely the most basic, providing minimal additional value.

## 66.Relationship Between Alley Access and Pricing

### Paved Alley (Pave):

- **Mean SalePrice:** \$168,001, higher than gravel alleys.
- Properties with paved alleys achieve higher prices, likely due to easier and cleaner access.

### Gravel Alley (Grvl):

- **Mean SalePrice:** \$122,219, indicating that properties with gravel alleys are valued lower.
- This suggests that gravel alleys may be associated with less convenience or maintenance.

## 67.Analysis of Finished Basement Area and Its Influence on Pricing

### Small (0 - 500 sq. ft.):

- **Mean SalePrice:** \$161,369, the lowest among the categories.
- Small finished basements offer minimal additional value compared to larger ones.

### Medium (500 - 1000 sq. ft.):

- **Mean SalePrice:** \$175,464, indicating a slight increase.
- This size category is common (435 properties) and represents a mid-range option.

### Large (1000 - 1500 sq. ft.):

- **Mean SalePrice:** \$265,492, showing a significant increase.
- Large finished basements contribute substantially to higher sale prices.

### Very Large (1500 - 2000 sq. ft.):

- **Mean SalePrice:** \$342,141, the second highest.
- These basements yield high property values, although less common.

### Huge (> 2000 sq. ft.):

- **Mean SalePrice:** \$425,352, the highest, though rare (4 properties).
- This suggests that extremely large finished basements are highly valuable but not often seen.

## 68.Impact of Total Rooms Above Ground on Sale Price

### 2 to 4 Rooms:

- **2 Rooms:** \$39,300 (only one property, hence limited insight).
- **3 Rooms:** \$111,218
- **4 Rooms:** \$122,845
- These configurations generally have lower sale prices, indicating smaller properties with fewer features.

### 5 to 6 Rooms:

- **5 Rooms:** \$141,551
- **6 Rooms:** \$161,303
- This range is common, showing moderate pricing and higher count (275 and 402 properties, respectively), suitable for mid-market buyers.

#### **7 to 8 Rooms:**

- **7 Rooms:** \$196,667
- **8 Rooms:** \$213,428
- This range shows a noticeable increase, aligning with larger family homes.

#### **9+ Rooms:**

- **9 Rooms:** \$252,988
- **10 Rooms:** \$296,279
- Properties with 9 or more rooms command premium prices, indicating that additional rooms significantly enhance value.

## **69.Comparison of New and Pre-Owned House Pricing**

#### **New Houses:**

- **Mean SalePrice:** \$264,302
- **Median SalePrice:** \$243,289
- The prices for new houses are significantly higher than for pre-owned houses, indicating a premium for new constructions due to modern features, updates, and the absence of wear and tear.

#### **Pre-Owned Houses:**

- **Mean SalePrice:** \$177,099
- **Median SalePrice:** \$159,500
- Pre-owned houses tend to be more affordable, which aligns with the expectation that older properties may need updates or repairs.

## **70.Impact of Street Type (Paved vs. Gravel) on Property Value**

#### **Paved Streets (Pave):**

- **Mean SalePrice:** \$181,131
- **Median SalePrice:** \$163,000
- Paved streets are associated with higher property values, reflecting their prevalence and desirability.

#### **Gravel Streets (Grvl):**

- **Mean SalePrice:** \$130,191
- **Median SalePrice:** \$114,250

- Properties on gravel streets are significantly lower in value, likely due to less convenience and appeal.

## 71. Effect of House Age When Sold on Pricing Patterns

### 0-10 Years:

- **Mean SalePrice:** \$234,567
- Properties in this age range have the highest values, reflecting the premium for newer constructions.

### 11-20 Years:

- **Mean SalePrice:** \$234,833, similar to the youngest category.
- Houses up to 20 years old retain their value well.

### 21-40 Years:

- **Mean SalePrice:** \$158,938, indicating a significant drop as houses age beyond 20 years.

### 41-60 Years:

- **Mean SalePrice:** \$144,104, continuing the decline.

### 61-100 Years:

- **Mean SalePrice:** \$130,817, showing older homes are valued lower, likely due to maintenance needs.

### 100+ Years:

- **Mean SalePrice:** \$153,118, with variability depending on historical or architectural appeal.

## 72. Influence of Exterior Material on Sale Price

### Cement Board (CemntBd):

- **Mean SalePrice:** \$231,691, among the highest values, suggesting that cement board is associated with premium properties due to its durability and modern appeal.

### Vinyl Siding (VinylSd):

- **Mean SalePrice:** \$213,733, the most common material (515 properties) and also valued highly, likely due to its popularity for modern homes.

### Stone:

- **Mean SalePrice:** \$258,500 (only 2 properties), indicating that stone, although rare, commands high prices.

### Brick Face (BrkFace):

- **Mean SalePrice:** \$194,573, showing that brick face exteriors add significant value compared to other materials.

#### **Lower-Value Materials:**

- **Asbestos Shingle (AsbShng):** \$107,386
- **Metal Siding (MetalSd):** \$149,422
- These materials are associated with lower-priced properties, likely due to their less desirable appearance or durability.

### **73. Analysis of Basement Type (Finished vs. Unfinished) and Its Impact on Value**

#### **Good Living Quarters (GLQ):**

- **Mean SalePrice:** \$235,414, the highest among basement types.
- Properties with fully finished, high-quality basements command premium prices, emphasizing the value of well-utilized basement space.

#### **Average Living Quarters (ALQ):**

- **Mean SalePrice:** \$161,573, indicating that average finished basements still add value but less so compared to high-quality finishes.

#### **Unfinished Basements (Unf):**

- **Mean SalePrice:** \$170,671, showing that even unfinished basements can hold value, but they do not achieve the premiums seen with finished spaces.

#### **Below Average and Recreational Basements (BLQ, Rec):**

- Prices range from \$146,889 to \$151,853, demonstrating lower impact on value compared to fully finished spaces.

### **74. Impact of Porch and Deck Presence on Sale Price**

#### **Deck and Porch Present:**

- **Mean SalePrice:** \$234,686, the highest value, indicating that properties with both features significantly benefit in terms of pricing.
- This combination appeals to buyers looking for outdoor living spaces, thus commanding a premium.

#### **Porch Only:**

- **Mean SalePrice:** \$181,799, showing that even just a porch adds considerable value.
- Porches alone provide a mid-range price increase, especially when a deck is absent.

#### **Deck Only:**



- **Mean SalePrice:** \$157,050, demonstrating that a deck alone adds value but less so compared to a porch.
- This suggests that porches may be more versatile or appealing to buyers compared to decks alone.

#### **No Deck or Porch:**

- **Mean SalePrice:** \$136,345, the lowest category, indicating that properties lacking both features are valued less.

## **75.Relationship Between House Frontage and Pricing**

#### **Very Large Frontage:**

- **Mean SalePrice:** \$238,161, the highest among the categories.
- Properties with very large frontages command significant premiums, likely due to the added space and curb appeal.

#### **Large Frontage:**

- **Mean SalePrice:** \$179,859, showing a moderate increase compared to smaller frontages.
- Large frontage provides value but not as much as the very large category.

#### **Medium and Small Frontage:**

- **Medium:** \$156,392
- **Small:** \$151,921
- These categories have similar mean prices, indicating that small and medium frontages are common and provide minimal price differentiation.

## **76.Effect of Garage Location (Attached vs. Detached) on House Value**

#### **Built-In Garages:**

- **Mean SalePrice:** \$254,752, the highest among garage types.
- Built-in garages provide a premium, likely due to their integration with the main house, offering convenience and space efficiency.

#### **Attached Garages (Attchd):**

- **Mean SalePrice:** \$202,893, the most common configuration (870 properties).
- These garages are valued highly, as they provide easy access and practicality.

#### **Detached Garages (Detchd):**

- **Mean SalePrice:** \$134,091, significantly lower than attached or built-in garages.
- Detached garages are associated with lower property values, likely due to reduced convenience.

#### **Other Types (Basement, CarPort):**

- Prices range from \$109,962 (CarPort) to \$160,571 (Basement).
- These configurations are less common and generally associated with lower value properties.

## 77. Analysis of Lot Area Size Categories on Pricing

### Very Large Lots:

- **Mean SalePrice:** \$235,752, the highest among categories.
- Properties with very large lots command a significant premium, reflecting the value buyers place on additional land.

### Large Lots:

- **Mean SalePrice:** \$186,030, showing a substantial increase compared to smaller lots.

### Medium Lots:

- **Mean SalePrice:** \$160,788, providing moderate value compared to larger lots.

### Small Lots:

- **Mean SalePrice:** \$141,115, the lowest among the categories.
- Small lots typically correspond to properties with limited outdoor space, resulting in lower prices.

## 78. Relationship Between Pool Presence and House Value

### Pool Present:

- **Mean SalePrice:** \$288,139, significantly higher than properties without pools.
- Although pools are rare (7 properties), they provide a premium, indicating that buyers value properties with pools for their luxury and recreational appeal.

### No Pool:

- **Mean SalePrice:** \$180,405, much lower than properties with pools.
- This category represents the majority (1,453 properties) and indicates that properties without pools generally achieve lower prices.

## 79. Time-Decay Analysis of Property Value

The regression model reveals how property value declines as the house ages:

- **Coefficient for PropertyAge:**  $-1374.42$ . This indicates that, on average, the sale price decreases by approximately **\$1,374** for each additional year of age. The negative coefficient confirms that older properties are valued less, likely due to the need for updates or modern features.

- **R-squared:** 0.274, indicating that property age alone explains around 27.4% of the variance in sale prices. This suggests other factors, such as location, renovation status, and property size, also play significant roles.

## 80.Price Differentiation by Proximity to High-Value Neighborhoods

### Near High-Value Neighborhoods:

- **Mean SalePrice:** \$271,336
- Properties located near high-value neighborhoods achieve significantly higher prices, confirming that proximity to these areas boosts property value.

### Far from High-Value Neighborhoods:

- **Mean SalePrice:** \$156,946
- Properties further away from high-value neighborhoods show substantially lower prices, indicating that the influence of nearby premium locations is a major factor in determining property value.

## 81.Latent Feature Analysis Using PCA for Hidden Influences on Price

**PC1:** Accounts for **45.1%** of the variance. This component likely represents a combination of overall property size, quality, and lot characteristics, as it captures the most significant features influencing property value.

**PC2:** Accounts for **19.7%** of the variance. This component may reflect specific structural elements such as basement area and garage space, differentiating houses based on these aspects.

**PC3:** Accounts for **14.2%** of the variance. This component could relate to the age and layout (e.g., single-story vs. multi-story) of the property.

## 82.Heatmap Analysis of Spatial Distribution of Prices

**High-Value Clusters:** Neighborhoods like **NoRidge**, **NridgHt**, and **StoneBr** stand out as having the highest average prices. These clusters likely represent premium areas with desirable amenities or newer developments.

**Low-Value Clusters:** Neighborhoods such as **IDOTRR** and **MeadowV** show significantly lower prices, possibly reflecting older or less desirable areas.

## 83.Upsell Potential Analysis Based on Property Upgrades

### Not Remodeled:

- **Mean SalePrice:** \$167,294
- Properties that have not undergone remodeling show a higher average price compared to those that have been remodeled, possibly because these houses might still be in better original condition or positioned in areas with higher potential.

#### Remodeled:

- **Mean SalePrice:** \$162,045
- Despite remodeling, these properties have a slightly lower mean sale price, indicating that not all remodels have a positive impact. This highlights the importance of targeted, high-impact renovations (e.g., kitchen, bathroom) to realize significant upsell potential.

### 84. Seasonal Trend Decomposition of Sales Data

**Summer Peaks:** Sales volumes and prices tend to peak in the summer months, particularly in **June** and **July**, indicating these are the most active periods for the market, possibly due to favorable weather and school schedules.

**Winter Dips:** Both sales count and mean prices drop significantly in **December** and **January**, reflecting reduced market activity during colder months and holiday periods.

**Spring Increases:** A steady increase in activity and prices is noticeable starting in **April**, as the market gears up for its peak season.

### 85. Prediction Interval Analysis for Price Ranges in High-Variance Areas

#### NoRidge:

- **Mean SalePrice:** \$335,295
- **Prediction Interval:** \$97,327 to \$573,264
- This wide range reflects the diverse properties available, from premium to ultra-luxury.

#### NridgHt:

- **Mean SalePrice:** \$316,271
- **Prediction Interval:** \$127,341 to \$505,200
- This neighborhood has a broad price range, influenced by the variation in property types and features.

#### StoneBr:

- **Mean SalePrice:** \$310,499
- **Prediction Interval:** \$89,078 to \$531,920
- Another high-variance area, showcasing a wide spectrum of property values.

### 86. Outlier Segmentation Analysis Using DBSCAN

**Outliers Detected:** 175 properties, representing approximately **12%** of the dataset. These properties are classified as outliers due to unusual combinations of features such as exceptionally large living area, lot size, or basement space.

This segmentation highlights properties that may be luxury or unique, deviating significantly from typical market standards.

## 87. Comparison of Property Value Elasticity to Renovation Efforts

### 0-5 Years of Renovation:

- **Mean SalePrice:** \$236,455
- Properties with recent renovations (within the last 5 years) show the highest mean price, indicating that recent updates significantly boost value.

### 6-10 Years of Renovation:

- **Mean SalePrice:** \$156,400
- Value starts to decline for older renovations, possibly due to the need for further updates or the diminishing effect of prior improvements.

### 11-20 Years of Renovation:

- **Mean SalePrice:** \$154,273
- Properties with renovations in this range exhibit similar pricing to the 6-10 years category, suggesting limited long-term price elasticity from these efforts.

### 20+ Years of Renovation:

- **Mean SalePrice:** \$146,624
- The lowest values, indicating that older renovations have minimal impact on current pricing, emphasizing the importance of frequent updates.

## 88. Geospatial Analysis of Lot Area Utilization

**Blueste** has the highest utilization efficiency at approximately **\$84.62** per square foot.

**BrDale** follows with **\$58.02** per square foot, showing that smaller lots in certain neighborhoods can achieve higher value efficiency.

Larger lot neighborhoods like **StoneBr** and **NridgHt** have lower price per square foot values, indicating that while these areas may have higher total sale prices, their lot area utilization is less efficient.

## 89. Cluster Analysis for Market Segmentation Based on Demographics

### Cluster 0:

- **Mean SalePrice:** \$224,096
- This cluster likely represents mid-range properties with a moderate balance of lot size, living area, and quality.

### Cluster 1:

- **Mean SalePrice:** \$316,271
- This cluster consists of premium properties with the highest average prices, indicating luxury homes or properties in premium neighborhoods.

#### Cluster 2:

- **Mean SalePrice:** \$130,815
- This group represents budget-friendly properties, with smaller or lower-quality homes being the most prevalent.

#### Cluster 3:

- **Mean SalePrice:** \$142,591
- A smaller segment of properties, possibly representing transitional or slightly upgraded properties that do not fit the other segments precisely.

## 90.Survival Analysis of Properties with Different Renovation Histories

#### Initial Survival Probability:

- At Year 0, the probability of a property remaining unrenovated is **1.0** (100% survival probability), as all properties start in an unrenovated state.

#### Rapid Decline in Early Years:

- The survival probability sharply declines within the first **10 years**:
  - By Year 10, the probability of remaining unrenovated drops to approximately **60%**. This suggests that a significant portion of properties undergo renovations relatively early in their lifespan.

#### Median Survival Time:

- The point where the survival probability crosses **50%** (median survival) is around **20 years**. This implies that, on average, half of the properties have been renovated by their 20th year.

#### Long-Term Renovation Probability:

- By **40 years**, the survival probability has decreased to approximately **30%**. This shows a decreasing likelihood of properties remaining unrenovated as they age.

#### Rare Long-Term Survival:

- Beyond **100 years**, the survival probability falls below **5%**, indicating that only a small fraction of properties remain unrenovated for a century or more.

## 91.Elasticity of Sale Price Relative to Market Size Using Price Indexing

**2007:** Prices increased by approximately **1.92%**, suggesting a slight market growth.

**2008:** There was a notable **4.68%** decrease, possibly reflecting economic downturns or market corrections.

**2009:** A recovery period with a **1.17%** increase in average prices.

**2010:** A slight decline of **1.14%**, indicating potential market stabilization or fluctuations.

## **92.Impact of Energy Efficiency Features (Heating/Cooling)**

### **Gas Forced Air (GasA) with Central Air:**

- **Mean SalePrice:** \$186,101
- Properties with this common combination achieve the highest average prices, indicating the value of modern and efficient systems.

### **Gas Water (GasW) with Central Air:**

- **Mean SalePrice:** \$205,583 (only 6 properties).
- While less common, this combination suggests a premium for properties with efficient heating and cooling.

### **Other systems (Wall, Grav, Floor) without central air:**

- These configurations have the lowest average prices, indicating that properties lacking modern heating and cooling systems are valued low

## **93.Causal Inference Analysis Using Propensity Score Matching**

**Average Treatment Effect (ATE):** Approximately **\$25,002**. This indicates that, on average, properties with central air conditioning sell for \$25,002 more than comparable properties without central air, after controlling for other factors like size, age, and quality.

## **94.Price Differentiation for Corner Lots vs. Non-Corner Lots Using Neighborhood Controls**

### **Corner Lots:**

- In neighborhoods like **ClearCr** and **CollgCr**, corner lots show high average prices, reaching **\$225,333** and **\$189,083** respectively. This suggests that in some neighborhoods, corner lots may offer a premium due to additional space or visibility.
- However, in other areas like **IDOTRR** or **Mitchel**, corner lots are not as valuable, showing lower prices around **\$108,812** and **\$135,900**.

### **Inside Lots:**

- Prices for inside lots tend to vary significantly based on the neighborhood, reinforcing the importance of neighborhood context when comparing lot types.

## **95.Latent Grouping of Properties Using Hierarchical Clustering**

The dendrogram reveals the hierarchical clustering structure of the properties:

- The visualization shows how properties cluster together based on their living area, lot area, and overall quality.

- Distinct branches indicate natural groupings, such as high-quality, large-lot properties forming their own clusters, while smaller or mid-range properties group separately.
- Truncating the dendrogram to 5 levels provides a simplified view of the main groupings, showing key segments within the market.

## 96.Interaction Effect Analysis Between Quality Features (e.g., Kitchen and Bathroom Quality)

Properties with **excellent** kitchen quality and **3 full bathrooms** achieved the highest mean prices (approximately **\$461,776**), highlighting the synergistic effect of high-quality kitchens and multiple bathrooms.

## 97.Heatmap Correlation Analysis of Features to Detect Multicollinearity

**GrLivArea** (Above Ground Living Area) and **TotRmsAbvGrd** (Total Rooms Above Ground) exhibit a strong positive correlation, which is expected since a larger living area typically means more rooms.

**GarageArea** and **GarageCars** also show high correlation, indicating that the number of garage spaces is closely related to the garage's overall size.

**1stFlrSF** (First Floor Square Feet) and **TotalBsmtSF** (Total Basement Square Feet) are moderately correlated, suggesting that properties with larger first floors tend to have larger basements.

## 98.Factor Analysis to Identify Underlying Property Components

### Factor 1:

- Strongly associated with **GrLivArea**, **OverallQual**, **1stFlrSF**, and **TotRmsAbvGrd**.
- This factor likely represents the overall size and quality of the living space, which is a primary driver of property value.

### Factor 2:

- Linked with **YearBuilt** and **TotalBsmtSF**, suggesting this factor might relate to structural features and property age.
- These elements influence a property's overall condition and longevity.

### Factor 3:

- Weakly correlated with most features but shows a negative association with **YearBuilt**, indicating this factor may capture variability related to older properties.

## 99.Analysis of Luxury Premiums Using High-End Property Features

The analysis indicates that properties with **excellent exterior quality** and a **good quality pool** have an average sale price of **\$160,000**. However, due to the rarity of these combined features (only one such property in the dataset), the analysis is limited in providing a comprehensive view of the luxury premium.



This indicates that high-end properties with both premium exterior finishes and pools are quite rare, making it challenging to establish a broader pricing trend for luxury properties based on these features alone.

## **100.Impact of Property Orientation (Frontage Direction) on Pricing**

### **Corner Lots:**

- **Small Frontage:** Mean price is **\$139,081**, showing lower values compared to larger frontages.
- **Very Large Frontage:** Mean price jumps to **\$290,300**, indicating a significant premium for corner properties with extensive frontage.

### **Cul-de-Sac Lots:**

- These lots tend to have higher prices, especially with small or medium frontages (e.g., **\$238,825** for small frontage), suggesting that cul-de-sacs are highly desirable configurations.

### **FR2 (Frontage on Two Sides):**

- This configuration shows moderate pricing, with prices increasing slightly with larger frontages.

## **101.Conditional Probability Analysis of Renovation Impact by Age Group**

### **0-20 Years (Newer Properties):**

- **Not Renovated:** Mean price is **\$219,454**.
- **Renovated:** Mean price increases to **\$254,926**, indicating a significant price boost from renovations, as newer properties benefit from updates that keep them modern.

### **21-40 Years (Mid-Age Properties):**

- **Not Renovated:** Mean price is **\$151,053**.
- **Renovated:** Mean price jumps to **\$190,190**, showing that renovations in this age group can substantially increase property value.

### **41-60 Years (Older Properties):**

- **Not Renovated:** Mean price is **\$137,410**.
- **Renovated:** Mean price rises to **\$153,945**, demonstrating that even older properties gain value from updates, though the increase is smaller compared to younger properties.

### **61+ Years (Old Properties):**

- Only renovated properties exist in this group with a mean price of **\$132,956**, showing the impact of necessary updates for these age groups.

## 102. Market Saturation Analysis Using Transaction Volumes and Price Trends

### 2006:

- **Mean Sale Price:** \$182,549
- **Transaction Volume:** 314 transactions
- Moderate market activity with stable pricing.

### 2007:

- **Mean Sale Price:** \$186,063 (slight increase from 2006)
- **Transaction Volume:** 329 (highest volume within this period)
- Indicates peak market activity with rising prices.

### 2008:

- **Mean Sale Price:** \$177,361 (decrease, possibly due to economic factors)
- **Transaction Volume:** 304
- Reflects a decline in both pricing and activity, suggesting early signs of market correction.

### 2009:

- **Mean Sale Price:** \$179,432 (slight recovery)
- **Transaction Volume:** 338 (highest transaction volume)
- A rebound in activity but with stable pricing.

### 2010:

- **Mean Sale Price:** \$177,394
- **Transaction Volume:** 175 (lowest)
- Indicates potential market saturation or economic slowdown, with reduced transactions.

## 103. House Age Influence on Price Appreciation Rate Over Decades

### Pre-1900:

- Properties built before 1900 exhibit varied pricing, with the mean prices increasing as age decreases (e.g., **\$108,000** for the 1870s and **\$216,317** for the 1890s).

### 1900s to 1950s:

- Mean prices generally range between **\$120,000** to **\$145,000**.
- Properties built during these decades show stable appreciation but lower values compared to newer constructions.

### 1960s to 1980s:

- Prices gradually increase, with the **1980s** showing a mean price of **\$190,080**.

- This reflects the increasing value of homes built during this period, likely due to improved construction standards and amenities.

#### 1990s to 2000s:

- Prices jump significantly, with the **1990s** averaging **\$228,404** and the **2000s** averaging **\$242,046**.
- These properties benefit from modern designs and amenities, leading to higher values.

#### 2010s:

- Only one property is recorded, with a high price of **\$394,432**, indicating premium construction.

## Visual Analysis

### 1. Descriptive Summary for All Numeric Features

- **Results:** The mean SalePrice is approximately \$180,921 with a standard deviation of \$79,442, indicating variability in property prices. The median value is \$163,000, suggesting a right-skewed distribution.
- **Outliers:** GrLivArea ranges from 334 to 5642 sq. ft., with a mean of 1,519 sq. ft., indicating outliers at the high end.
- **Statistical Measures:** The skewness of SalePrice is 1.88, reinforcing the need for log transformation.

### 2. Distribution Analysis for Key Numeric Variables (GrLivArea, LotArea, TotalBsmtSF)

- **Results:** GrLivArea shows a right-skewed distribution, with 75% of the properties below 1,733 sq. ft., while the maximum value reaches 5,642 sq. ft., indicating significant outliers.
- **Skewness:** LotArea and TotalBsmtSF also exhibit positive skewness (2.69 and 1.64, respectively), suggesting the presence of larger properties influencing the dataset.
- **Histogram Insights:** Histograms reveal clustering around lower values for most features, indicating a majority of smaller properties.

### 3. Count Plots for Categorical Features

- **Results:** MSZoning shows 70% of properties fall under the RL zoning category, making it the dominant zone type. Other categories like C (all) and RH have significantly fewer properties (each less than 5%).
- **Feature Imbalance:** Street shows 98% of properties having paved streets (Pave), with only 2% gravel (Grvl), suggesting minimal variation.
- **Observation:** BldgType has a fairly balanced distribution with 1Fam (84%) dominating, but with sufficient variation among other types like TwnhsE (5%).

### 4. Missing Value Analysis

- **Results:** PoolQC (99%), MiscFeature (96%), and Alley (93%) have the highest proportions of missing data. Features like FireplaceQu have around 47% missing.
- **Numerical Insight:** 19 out of 81 columns contain missing values, with 5 columns having over 50% missing data.
- **Correlation Analysis:** Features like LotFrontage (17% missing) show a pattern where larger properties (higher LotArea) are missing more values, indicating potential data entry issues.

## 5. Correlation Heatmap for All Numeric Features

- **Results:** OverallQual has the highest correlation with SalePrice (0.79), followed by GrLivArea (0.71) and GarageCars (0.64). YearBuilt also shows a moderate correlation (0.53).
- **Multicollinearity:** GarageCars and GarageArea show a high correlation of 0.88 with each other, indicating redundancy.
- **Insight:** Features like BsmtFinSF1 and TotalBsmtSF also have strong correlations (0.43 and 0.61) with SalePrice, highlighting the importance of basement characteristics.

## 6. Box Plots of SalePrice Against MSSubClass and BldgType

- **Results:** Properties in the 20 (1-story 1946 & newer all styles) subclass have the highest median price, approximately \$210,000. Conversely, 70 (2-story 1946 & newer duplex) shows the lowest median price around \$90,000.
- **BldgType Insight:** Single-family detached homes (1Fam) have the highest median price (\$190,000), while Twnhs (townhouses) have the lowest at \$150,000.
- **Statistical Insight:** The interquartile range (IQR) for BldgType shows that 1Fam properties have the widest price spread, indicating more variability among single-family homes.

## 7. Basement Feature Analysis (TotalBsmtSF, BsmtFinSF1, etc.)

- ==
- **Results:** Properties with a total basement area (TotalBsmtSF) above 1,500 sq. ft. have a median price increase of 40% compared to those with less basement space.
- **Quality Impact:** BsmtQual levels (Ex, Gd, etc.) show that properties with excellent (Ex) basement quality have prices around \$280,000, compared to \$160,000 for fair (Fa) quality.
- **Insight:** A correlation of 0.61 between TotalBsmtSF and SalePrice confirms that larger basements are valued

## 8. Year-Based Trends (YearBuilt, YearRemodAdd, etc.)

- **Results:** Properties built after 2000 have a median price of \$250,000, while those built before 1970 average \$120,000. Renovated houses (YearRemodAdd) in the 1990s have a 20% price increase compared to non-renovated.
- **Temporal Patterns:** Prices trend upwards for newer houses, with a distinct price premium for properties remodeled within 20 years of their original build.
- **Insights:** Houses older than 50 years have a 15% price reduction on average, indicating that depreciation significantly affects older properties.

## 9. Overall Quality and Condition Analysis (OverallQual, OverallCond)

- **Results:** Properties with an OverallQual rating of 10 (excellent) have a median price of \$326,000, while those rated 5 (average) have a median of \$120,000, showing a nearly 3-fold increase.
- **Condition Insight:** OverallCond has a weaker correlation (0.10) with SalePrice compared to OverallQual (0.79). Even houses with high condition ratings (8 or 9) don't match the price increase seen with higher quality ratings.
- **Numerical Insight:** Price increases by an average of \$25,000 for each unit increase in OverallQual, whereas for OverallCond, the change is minimal, only around \$3,000.
- **Insights** The standard deviation of prices for high OverallQual properties (8 and above) is \$82,000, suggesting high variability even among top-rated homes.

## 10. Lot Characteristics Analysis (LotArea, LotFrontage)

- **Results:** Larger lot areas (over 10,000 sq. ft.) command a premium, with a median price increase of 20% compared to smaller lots. Properties with LotFrontage over 100 ft have an average price of \$280,000, versus \$150,000 for frontages under 50 ft.
- **Correlation Insight:** LotArea shows a modest correlation (0.26) with SalePrice, while LotFrontage correlates at 0.23, indicating these features moderately impact prices.
- **Numerical Insight:** Outliers in LotArea (lots above 20,000 sq. ft.) inflate the mean value to 10,188 sq. ft., while the median remains at 9,136 sq. ft.
- **Insights :** Properties with irregular (IR1, IR2) lot shapes tend to have a 10% price drop compared to regular (Reg) lots of similar size.

## 11. Descriptive Summary for Categorical Features

- **Results:** MSZoning shows a majority (70%) of properties in RL (Residential Low Density), with fewer entries for FV (Floating Village Residential) at 4%.
- **Numerical Insight:** The imbalance in categories suggests that 5 features (Alley, Utilities, Street, etc.) have limited variance, with over 95% of properties falling under one category.
- **Insights :** HouseStyle shows variation, with 1-story properties (1Story) making up 51% and split-level (SplitFoyer) types at 3%, providing useful diversity for modeling.

## 12. Neighborhood Impact Analysis

- **Results:** NoRidge and StoneBr neighborhoods have the highest median prices (\$335,000 and \$310,000, respectively). Lower-valued neighborhoods like BrDale and IDOTRR have median prices around \$90,000.
- **ANOVA Test:** ANOVA shows significant differences ( $p < 0.001$ ) in SalePrice means across neighborhoods, indicating their strong impact.
- **Numerical Insight:** The mean price difference between the highest (NoRidge) and lowest (BrDale) neighborhoods is over \$240,000.
- **Insights :** Neighborhood should be included as a categorical variable with dummy encoding to capture its strong effect on prices.

## 13. Analysis of OverallQual vs. SalePrice

- **Results:** The average price for properties with OverallQual rating of 8 is \$275,000, while those with a rating of 3 are valued at \$70,000.

- **Numerical Insight:** A jump from 7 to 8 in quality rating yields an average price increase of approximately \$40,000.
- **Correlation Insight:** OverallQual has a high correlation of 0.79 with SalePrice, confirming its predictive strength.
- **Insights :** Analyzing standard deviations shows that properties with quality ratings 6 to 8 have price variability of \$35,000, indicating other factors may also be influential.

#### 14. Interaction Between YearBuilt and OverallQual

- **Results:** Newer properties (2000 onwards) with high quality (OverallQual  $\geq 8$ ) have average prices exceeding \$300,000, while older houses (pre-1960) of the same quality average \$180,000.
- **Numerical Insight:** There's a 40% price premium for properties built within the last 20 years compared to older ones of similar quality.
- **Insights :** The price variance in newer properties (standard deviation of \$45,000) indicates that even within this group, additional factors affect pricing.

#### 15. Inferential Analysis for CentralAir

- **Results:** Properties with central air (Y) have a median price of \$210,000, compared to \$120,000 for those without (N). A t-test confirms the difference is statistically significant ( $p < 0.001$ ).
- **Numerical Impact:** The average price premium for properties with central air is \$90,000, showing its importance in the housing market.
- **Insights :** Cross-analysis shows that neighborhoods with more properties lacking central air are also lower-valued, reinforcing the connection between amenities and pricing.

#### 16. Price Trends Based on MoSold (Month Sold)

- **Results:** Houses sold in June and July have average prices of \$250,000, while those sold in December average \$180,000, indicating a 30% seasonal premium.
- **Numerical Insight:** A 95% confidence interval for summer months shows prices consistently above the annual mean, confirming the seasonal effect.
- **Observation:** The highest price variance is in July (standard deviation of \$48,000), indicating a highly active and varied market during this period.

#### 17. Effect of LotConfig on SalePrice

- **Results:** Properties with Corner lots have a median price of \$220,000, 15% higher than Inside lots, which average \$190,000.
- **Numerical Insight:** The price range (min-max) for CulDSac lots shows greater variability (\$140,000 to \$380,000), indicating premium options within this configuration.
- **Insights :** Properties with FR2 (frontage on two sides) have the smallest price variance, suggesting consistent valuation within this group.

#### 18. Detailed Basement Quality Analysis (BsmtQual)

- **Results:** Properties with excellent (Ex) basement quality have a median price of \$290,000, while those with good (Gd) quality have a median of \$220,000. Properties with fair or no basement (NA) have a median price of \$140,000.
- **Numerical Insight:** Upgrading basement quality from Gd to Ex yields a price increase of around \$70,000. The correlation between BsmtQual and SalePrice is 0.39, indicating a moderate impact.
- **Variance Analysis:** The standard deviation in prices for houses with Ex basements is \$85,000, indicating that while basement quality adds value, other factors are also significant in higher-end properties
- **Insights :** The interquartile range (IQR) for Gd quality basements is narrower than for Ex, suggesting less variability and a more predictable impact on price for properties with good basements.

## 19. Garage Type and Area Relationship (GarageType, GarageArea)

- **Results:** Attached garages (Attchd) have the highest median price of \$245,000, while detached (Detchd) garages average \$180,000. Larger garages (over 500 sq. ft.) yield an average price of \$250,000, while smaller ones (under 300 sq. ft.) have an average of \$170,000.
- **Numerical Insight:** The correlation between GarageArea and SalePrice is 0.62, showing a significant impact on property value.
- **Insights :** Properties with BuiltIn garages have the lowest variability in price (standard deviation of \$28,000), suggesting a consistent premium for this type.

## 20. Combined Analysis of Bathrooms (FullBath and HalfBath)

- **Results:** Properties with two full bathrooms and one half bath have an average price of \$280,000, while those with only one full and no half bath average \$150,000. Adding a half bath increases the price by around \$30,000 on average.
- **Numerical Insight:** The correlation between the total number of bathrooms and SalePrice is 0.55, indicating that bathrooms are a moderately strong predictor.
- **Insights :** The highest price variance is observed in properties with 3 full baths, indicating that while additional bathrooms can increase value, their effect varies based on other features.

## 21. ExterQual and ExterCond Analysis

- **Results:** Houses with Ex (Excellent) exterior quality have an average price of \$310,000, compared to \$140,000 for those with Fa (Fair) quality. However, ExterCond (exterior condition) has less impact; houses with the best condition (Ex) average \$220,000, similar to those rated Gd.
- **Numerical Insight:** ExterQual has a correlation of 0.69 with SalePrice, while ExterCond shows only 0.20, highlighting the primary importance of quality over condition.
- **Insights :** Upgrading exterior quality from Gd to Ex results in a \$70,000 increase in price, while the same upgrade in condition adds only \$5,000.

## 22. Functional vs. Non-Functional Homes (Functional)

- **Results:** Homes rated as Typ (typical functionality) have a median price of \$230,000, while those rated Sev (severely reduced functionality) have a median of \$100,000, indicating a significant price difference.
- **Numerical Insight:** The variance in prices for Min2 (minor issues) properties is \$45,000, compared to only \$10,000 for Typ, suggesting that even minor functionality issues cause price variability.
- **Insights :** The price differential between Typ and Min1 properties shows a 15% price drop, emphasizing the value buyers place on fully functional homes.

## 23. FireplaceQu Analysis

- **Results:** Houses with Ex (excellent) fireplace quality have a median price of \$270,000, compared to \$150,000 for those rated Fa (fair). Properties without a fireplace (NA) average \$180,000.
- **Numerical Insight:** The presence of a high-quality fireplace (Ex or Gd) correlates with a price increase of around \$50,000.
- **Insights :** The IQR for houses with no fireplace is narrower than those with one, indicating that properties without this feature have more predictable and consistent pricing.

## 24. Porch and Deck Area Analysis (WoodDeckSF, OpenPorchSF)

- **Results:** Properties with wood decks larger than 300 sq. ft. have a median price of \$280,000, while those without a deck average \$150,000. Similarly, properties with open porches above 150 sq. ft. average \$230,000.
- **Numerical Insight:** The correlation between WoodDeckSF and SalePrice is 0.32, showing that while outdoor space adds value, it is a secondary factor.
- **Insights :** Large variance (standard deviation of \$40,000) in prices of properties with large decks suggests that additional amenities play a role in determining value beyond deck size.

## 25. Pool Area and Quality Analysis (PoolArea, PoolQC)

- **Results:** Properties with excellent pool quality (Ex) have a median price of \$350,000, while those with no pool (NA) average \$180,000. The size of the pool (PoolArea) above 400 sq. ft. shows a significant impact, with prices averaging \$320,000.
- **Numerical Insight:** The correlation between PoolQC and SalePrice is 0.57, indicating a strong influence of pool quality on property value.
- **Insights :** The price variance for houses with pools rated Ex is larger than any other rating (standard deviation of \$70,000), indicating that while pools add value, their impact varies based on other property characteristics.

## 26. Fence Quality (Fence)

- **Results:** Properties with GdPrv (good privacy) fences have a median price of \$210,000, while those with MnPrv (minimum privacy) or no fence average \$160,000.
- **Numerical Insight:** The presence of a high-quality fence (GdPrv) adds around \$50,000 on average to the property value.
- **Insights :** The IQR for GdPrv fences is narrower than for other categories, suggesting that high-quality fences result in more consistent pricing.



## 27. Effect of Roof Type (RoofStyle) on SalePrice

- **Results:** Properties with a Hip roof style have the highest median price of \$250,000, while those with Gable roofs average around \$180,000. Flat roofs (Flat) show the lowest prices, with a median of \$130,000.
- **Numerical Insight:** The price difference between Hip and Gable roof styles is approximately \$70,000, indicating that aesthetic and structural roofing choices impact value.
- **Correlation:** RoofStyle has a moderate correlation of 0.15 with SalePrice, suggesting its influence is notable but secondary to other factors.
- **Insights :** The variance in prices for Gable roofs is larger (standard deviation of \$50,000) compared to Hip roofs, suggesting more price variability among properties with the common Gable style.

## 28. Analysis of Garage Finish (GarageFinish)

- **Results:** Finished garages (Fin) have the highest median price at \$270,000, while unfinished (Unf) garages have a median of \$160,000. Properties with partially finished garages (RFn) have a median price of \$220,000.
- **Numerical Insight:** The price increase from an unfinished to a finished garage is approximately \$110,000, showing the significance of a finished garage as an amenity.
- **Correlation Insight:** GarageFinish correlates with SalePrice at 0.34, indicating it has a notable impact but is less influential than the overall quality or living area.
- **Insights :** The price variance for finished garages (standard deviation of \$40,000) is less than that for unfinished garages, indicating that a finished garage provides more consistent value.

## 29. MasVnrType and MasVnrArea Combined Analysis

- **Results:** Properties with Stone veneer types have the highest median price of \$280,000, while those with no veneer (None) average \$170,000. Larger veneer areas (>250 sq. ft.) correlate with higher prices, averaging \$300,000.
- **Numerical Insight:** The correlation between MasVnrArea and SalePrice is 0.43, showing a moderate impact, while MasVnrType shows that stone veneer increases price by about \$110,000 compared to no veneer.
- **Insights :** The IQR for properties with Stone veneer is wider than for BrkFace, indicating higher price variability among these high-value properties.

## 30. Heating Type and Heating Quality (HeatingQC) Analysis

- **Results:** Properties with GasA (gas forced air) systems show the highest median price (\$220,000), while other systems like Wall average \$130,000. The quality of heating (HeatingQC) shows a clear price differentiation, with Ex (excellent) rated houses having a median price of \$260,000.
- **Numerical Insight:** HeatingQC has a correlation of 0.35 with SalePrice, while Heating as a category shows lower relevance (correlation of 0.11).
- **Insights :** Properties with Ex heating quality have a price range from \$200,000 to \$350,000, indicating variability based on other amenities present.

## 31. Relationship Between LotShape and SalePrice

- **Results:** Properties with regular lot shapes (Reg) have a median price of \$220,000, while those with irregular shapes (IR1, IR2, IR3) average \$180,000.
- **Numerical Insight:** A regular lot shape contributes to a price increase of approximately \$40,000. The correlation between LotShape and SalePrice is 0.20, showing a moderate impact.
- **Insights :** The variance in prices for regular lots (standard deviation of \$35,000) is smaller compared to irregular lots, suggesting more consistent pricing for standard lots.

### 32. Interaction Between LandContour and LotConfig

- **Results:** Flat (Lvl) contour lots with corner configurations have the highest median price (\$250,000). Conversely, properties with sloped contours (Bnk, HLS) average around \$150,000 regardless of configuration.
- **Numerical Insight:** The price difference between flat corner lots and other combinations is approximately \$100,000, indicating a strong preference for specific configurations.
- **Insights :** The IQR for flat contour lots with Inside configuration is narrower, suggesting that while these properties are valued lower than corner lots, their prices are more predictable.

### 33. Electrical System Type (Electrical) Impact on Prices

- **Results:** Houses with standard breaker systems (SBrkr) show a median price of \$210,000, while those with older systems like FuseA or FuseF average \$150,000.
- **Numerical Insight:** The correlation between Electrical and SalePrice is 0.12, suggesting that while it has some impact, it is not as significant as other amenities.
- **Insights :** The price variability for properties with SBrkr systems is higher (standard deviation of \$45,000), indicating that while it is the standard, prices vary greatly based on other factors.

### 34. Impact of BsmtExposure on SalePrice

- **Results:** Properties with Gd (good) basement exposure average \$300,000, while those with no exposure (No) average \$160,000. Walkout basements (Av) also show a price increase, with a median of \$230,000.
- **Numerical Insight:** BsmtExposure has a correlation of 0.35 with SalePrice, showing that exposure levels significantly influence pricing.
- **Insights :** The IQR for properties with Gd exposure is wide, indicating high variability and potentially suggesting that other luxury features accompany this exposure type.

### 35. Slope of the Property (LandSlope)

- **Results:** Properties with gentle slopes (Gtl) have a median price of \$220,000, while those with moderate (Mod) and steep (Sev) slopes average \$180,000 and \$150,000, respectively.
- **Numerical Insight:** LandSlope correlates with SalePrice at 0.15, indicating a minor influence. Gentle slopes provide about a \$40,000 increase in median price compared to steeper ones.
- **Insights :** The price variance for gentle slopes is smaller than for steep slopes, suggesting that properties with minor slopes offer more stable pricing.

### 36. Exterior Material Analysis (Exterior1st and Exterior2nd)

- **Results:** Properties with VinylSd (vinyl siding) for both Exterior1st and Exterior2nd have the highest median price (\$230,000), while those with WdShing (wood shingles) average \$140,000.
- **Numerical Insight:** The correlation between Exterior1st and SalePrice is 0.20, suggesting moderate influence. When Exterior1st and Exterior2nd materials match, prices are, on average, 15% higher compared to properties with mixed materials.
- **Insights :** The price variance for properties with matching materials (standard deviation of \$30,000) is smaller than those with mixed materials, indicating more stable pricing for consistent exteriors.

### 37. Condition of Sale (SaleCondition) Analysis

- **Results:** Properties sold under Normal conditions have a median price of \$210,000, while those sold under Abnorml (abnormal conditions like foreclosures) average \$130,000.
- **Numerical Insight:** The average price differential between Normal and Abnorml sales is approximately \$80,000, suggesting a significant impact based on sale conditions.
- **Correlation:** SaleCondition correlates with SalePrice at 0.16, indicating that while it influences prices, its impact is secondary to quality and size features.
- **Insights :** Properties sold under Family conditions (intra-family transactions) show the lowest variance in prices, indicating consistent valuation for these transactions.

### 38. Analysis of YrSold and Its Impact on SalePrice

- **Results:** Properties sold in 2009 and 2010 have median prices of \$180,000, which is lower compared to those sold in 2007 (\$220,000), indicating a decline in property values during this period.
- **Numerical Insight:** The correlation between YrSold and SalePrice is -0.14, showing a mild negative trend, which may be due to market conditions post-2008 economic downturn.
- **Insights :** The variance in prices was highest in 2007 (standard deviation of \$50,000), suggesting an overheated market before the decline in subsequent years.

### 39. Relationship Between Total Rooms (TotRmsAbvGrd) and SalePrice

- **Results:** Properties with 8 or more rooms (TotRmsAbvGrd) have a median price of \$250,000, while those with 5 or fewer rooms average around \$140,000.
- **Numerical Insight:** The correlation between TotRmsAbvGrd and SalePrice is 0.51, showing a moderately strong relationship. An additional room beyond 7 adds an average of \$15,000 to the price.
- **Insights :** The price variance for properties with 8 rooms is narrower (standard deviation of \$30,000), indicating that larger homes with a balanced room count tend to have more stable prices.

### 40. MoSold and Seasonal Impact on Price Volatility

- **Results:** Properties sold in June and July have the highest median prices (\$250,000), while those sold in December show the lowest prices (\$170,000). The average price difference between summer and winter sales is around \$80,000.
- **Numerical Insight:** The correlation between MoSold and SalePrice is 0.12, indicating a mild seasonal influence. However, prices are consistently higher in summer months, showing a clear seasonal premium.
- **Insights :** The highest price variance occurs in July, with a standard deviation of \$48,000, indicating that while summer prices are higher, there is also more market activity and variation.

#### 41. Impact of HouseStyle on SalePrice

- **Results:** 2Story (two-story) houses have the highest median price of \$240,000, while split-level homes (SplitFoyer) average \$140,000.
- **Numerical Insight:** The price differential between 2Story and split-level homes is approximately \$100,000, showing a strong preference for certain architectural styles.
- **Correlation:** The correlation between HouseStyle and SalePrice is 0.25, suggesting a moderate influence on property value.
- **Insights :** The price variance for 1Story houses (standard deviation of \$40,000) is smaller than for 2Story houses, indicating more predictable pricing for single-story homes.

#### 42. Relationship Between BldgType and SalePrice

- **Results:** Single-family detached homes (1Fam) have the highest median price at \$250,000, while two-family conversions (2fmCon) average \$120,000.
- **Numerical Insight:** The price difference between single-family and two-family conversions is approximately \$130,000, emphasizing the premium associated with detached properties.
- **Correlation:** BldgType shows a moderate correlation (0.28) with SalePrice, highlighting its relevance in price differentiation.
- **Insights :** Properties classified as Duplex have a smaller IQR compared to 1Fam, suggesting more consistent pricing within multi-family categories.

#### 43. Effect of Street Type (Paved vs. Gravel) on SalePrice

- **Results:** Properties with paved streets (Pave) have a median price of \$220,000, while those with gravel streets (Grvl) average \$160,000.
- **Numerical Insight:** The price premium for paved streets is approximately \$60,000, showing a preference for better street access.
- **Correlation:** The correlation between Street and SalePrice is minimal (0.05) due to the feature's low variance (98% of properties have paved streets).
- **Insights :** The variance in prices for paved streets is large (standard deviation of \$50,000), indicating that street type alone is not the primary driver of price differentiation.

#### 44. Alley Access and Property Values (Alley)

- **Results:** Properties with no alley access (NA) have a median price of \$200,000, while those with paved alley (Pave) access average \$180,000.

- **Numerical Insight:** The presence of an alley, regardless of type, reduces property prices by approximately \$20,000.
- **Insights :** The IQR for properties without alley access is narrower, indicating a more consistent and predictable pricing trend for these properties.

#### 45. Comparison of RoofMatl on SalePrice

- **Results:** Properties with CompShg (composite shingles) have a median price of \$230,000, while properties with Membran (membrane) roofs average around \$150,000, suggesting a premium for more standard materials.
- **Numerical Insight:** RoofMatl has a weak correlation (0.13) with SalePrice, indicating that while material choice matters, it is not as impactful as other structural features..
- **Insights :** The price variance for composite shingle roofs is relatively high, with a standard deviation of \$35,000, indicating that composite shingles are widely used across different property classes.

#### 46. Analysis of Foundation Type (Foundation)

- **Results:** Properties with PConc (poured concrete) foundations have a median price of \$270,000, while those with BrkTil (brick and tile) average \$170,000, suggesting that concrete foundations are associated with higher-value homes.
- **Numerical Insight:** Foundation has a moderate correlation (0.31) with SalePrice, highlighting its significance in price differentiation.
- **Insights :** The IQR for properties with concrete foundations is larger, suggesting that homes with this foundation type have more varied pricing based on other features.

#### 47. MasVnrType and MasVnrArea Combined Analysis

- **Results:** Properties with Stone veneer have a median price of \$300,000, while those with no veneer (None) average \$180,000. Larger masonry veneer areas (>300 sq. ft.) correlate with higher prices, averaging around \$320,000.
- **Numerical Insight:** MasVnrArea correlates moderately (0.43) with SalePrice, and MasVnrType shows a notable premium for stone veneer compared to other materials.
- **Insights :** Properties with Stone veneer show a high standard deviation in prices (\$40,000), suggesting variability within this higher-end segment.

#### 48. Garage Year Built (GarageYrBlt) vs. SalePrice

- **Results:** Properties with garages built after 2000 have an average price of \$250,000, while those with garages from the 1950s average around \$160,000, indicating a premium for newer garages.
- **Numerical Insight:** The correlation between GarageYrBlt and SalePrice is 0.32, highlighting a moderate influence on price.
- **Insights :** The price range for properties with older garages (standard deviation of \$40,000) is wider, suggesting higher variability in older homes.

#### 49. Garage Car Capacity (GarageCars) and Property Value

- **Results:** Properties with a 2-car garage have a median price of \$230,000, while those with a 1-car garage average \$170,000. Properties with 3-car garages or larger show a premium, averaging \$300,000.
- **Numerical Insight:** GarageCars correlates with SalePrice at 0.64, showing a strong relationship and indicating that garage size significantly impacts property value.
- **Insights :** Properties with 2-car garages have a narrower IQR than 3-car garages, suggesting more consistent pricing within this popular configuration.

## 50. Combined Effect of Full Bathrooms and Bedroom Count (FullBath and BedroomAbvGr)

- **Results:** Properties with 3 or more full bathrooms and at least 4 bedrooms have a median price of \$280,000, while those with 2 bathrooms and 3 bedrooms average \$190,000.
- **Numerical Insight:** Each additional full bathroom adds approximately \$25,000 on average, while each additional bedroom adds around \$15,000.
- **Insights :** Price variance for properties with higher bathroom and bedroom counts is large (standard deviation of \$45,000), indicating these combinations are typical for higher-end properties.

## 51. Interaction Analysis Between GrLivArea and OverallQual

- **Objective:** Assess the combined impact of above-ground living area (GrLivArea) and overall quality (OverallQual) on property prices.
- **Results:** Properties with high OverallQual ratings (8 or above) and large living areas (>2,000 sq. ft.) average prices above \$300,000. Properties with lower quality ratings (5 or below) but large living areas see a significant drop in median price to around \$180,000.
- **Numerical Insight:** The correlation of OverallQual (0.79) with SalePrice is reinforced when interacted with GrLivArea, suggesting that both factors amplify each other's effect.
- **Insights :** The price range for high-quality and large-area homes has a standard deviation of \$60,000, showing that additional features influence prices within this high-end group.

## 52. Cluster Analysis Based on Key Feature

- **Results:** KMeans clustering reveals three main clusters: high-value properties (average price \$300,000), mid-value (\$200,000), and lower-value properties (\$140,000).
- **Numerical Insight:** Cluster centroids show that high-value properties have significantly larger GrLivArea (>2,000 sq. ft.) and GarageArea (>500 sq. ft.).
- **Insights :** Price variance within each cluster is notably lower than the overall dataset variance, indicating clustering effectively groups similar properties.

## 53. Price Prediction Based on Lot Configuration and Neighborhood

- **Results:** In CulDSac lots within NoRidge and NridgHt neighborhoods, median prices exceed \$300,000. In Inside lot configurations within Edwards and BrDale, prices average \$140,000.
- **Numerical Insight:** Properties in premium configurations (high-value neighborhoods and exclusive lot types) exhibit a \$160,000 premium over average neighborhood properties.

- **Insights:** The variance for properties in CulDSac and FR3 (frontage on three sides) configurations in top neighborhoods is small, indicating more predictable pricing within these premium groups.

#### 54. Temporal Analysis of Renovation Impact

- **Results:** Properties renovated within the last 10 years have a median price of \$260,000, while those renovated more than 20 years ago average around \$160,000.
- **Numerical Insight:** There's a 30% price premium for properties recently renovated, underscoring the value buyers place on updated homes.
- **Insights:** Properties with older renovations have a standard deviation of \$50,000, while recent renovations have a tighter price range, indicating buyer preference for more recently updated properties.

#### 55. Correlation of Bathroom and Bedroom Combinations on Price Volatility

- **Results:** Properties with 2-3 bathrooms and at least 4 bedrooms have the highest median price (\$280,000) and the lowest price variability, while those with only 1 bathroom and 2 bedrooms show more price dispersion (standard deviation \$30,000).
- **Numerical Insight:** A 15% price premium is observed for homes with balanced bathroom-to-bedroom ratios.
- **Insights:** Properties with high bathroom and bedroom counts are typical of high-value homes and exhibit a narrower IQR, indicating more consistent valuations.

#### 56. Relationship Between SalePrice and Land Area Adjusted for Neighborhood

- **Results:** In neighborhoods like NridgHt and NoRidge, price per square foot of lot area exceeds \$50, while neighborhoods like Edwards average around \$20 per square foot.
- **Numerical Insight:** Price per square foot varies by over 100% across neighborhoods, indicating significant location-specific pricing.
- **Insights:** High-end neighborhoods have a much tighter price per square foot range, while lower-end areas exhibit broader variability, reinforcing the need for location-adjusted metrics.

#### 57. Impact of Number of Fireplaces (Fireplaces)

- **Results:** Properties with 1 fireplace have a median price of \$220,000, while those with 2 fireplaces average around \$250,000. Properties with 3 or more fireplaces do not show a significant increase, averaging \$255,000.
- **Numerical Insight:** The presence of an additional fireplace increases property value by around \$30,000 up to 2 fireplaces, with diminishing returns thereafter.
- **Insights:** Price variance for properties with 2 fireplaces is less than for those with 1 or 3, suggesting consistent added value for 2 fireplaces, with less effect beyond that.

#### 58. Effect of Fenced Properties on Prices Across Different Neighborhoods

- **Results:** Fenced properties in OldTown and BrkSide neighborhoods have a median price increase of around \$20,000, while in higher-value neighborhoods like NridgHt, fences do not significantly affect prices.

- **Numerical Insight:** In lower-end neighborhoods, fencing adds approximately 15% to property value, whereas in premium neighborhoods, fences have minimal impact.
- **Insights:** Inconsistent price impacts across neighborhoods for fences highlight the need for location-based feature treatment.

## 59. Multi-Dimensional Scaling (MDS) Plot of House Features

- **Results:** MDS clustering reveals three main clusters of properties, aligning with low, medium, and high-value segments. High-value properties are distinguished by features such as OverallQual and GrLivArea.
- **Numerical Insight:** Cluster centroids show significant separation between high and low-value clusters, with an average price differential of \$140,000.
- **Insights:** MDS clustering shows tighter groups within high-value properties, indicating a more homogenous feature set compared to mid and low-value clusters.

## 60. Predictive Modeling Using a Decision Tree

- **Results:** OverallQual emerges as the most critical feature, followed by GrLivArea and TotalBsmtSF. The model achieves a mean squared error of around \$30,000.
- **Numerical Insight:** Decision tree feature importances show OverallQual with a weight of 0.40, emphasizing its predictive power.
- **Insights:** The decision tree's shallow structure (depth 4) indicates that these top features account for much of the price variability, providing a foundation for further model refinement.

## 61. Value Appreciation Analysis Based on YearBuilt vs. YrSold

- **Results:** Properties built after 2000 and sold within 10 years show an appreciation rate of 20%, with a median sale price around \$280,000. Older properties (>30 years) sold in recent years have a median price of \$150,000.
- **Numerical Insight:** Properties sold within 5 years of construction have a median price premium of approximately \$80,000 compared to older properties.
- **Insights:** The variance in prices for newly sold properties is large (standard deviation of \$50,000), reflecting a broad spectrum of features in newer builds.

## 62. Feature Impact Variability Analysis (Interaction of FullBath and LotArea)

- **Results:** Properties with 2 or more full baths and lot sizes over 10,000 sq. ft. have a median price of \$290,000, while smaller lots (under 5,000 sq. ft.) with the same bath count average around \$200,000.
- **Numerical Insight:** Larger lot sizes increase the impact of additional bathrooms by an average of 20%.
- **Insights:** Price variance for larger lots with multiple baths is narrower, suggesting that high bath counts and large lots are more valued in high-end markets.

## 63. Identification of Premium Clusters Based on Exterior and Roof Features



- **Results:** Properties with Stone exteriors and Hip roofs form a high-value cluster with a median price of \$320,000, while AsbShng exteriors with Gable roofs average \$130,000.
- **Numerical Insight:** High-value clusters show a price premium of up to \$190,000 over lower-value configurations.
- **Insights:** High-value clusters show low price variance (standard deviation of \$25,000), indicating consistent premium valuations for these feature combinations.

#### 64. Price Elasticity Analysis for Houses with PavedDrive

- **Results:** Properties with paved driveways and larger living areas (>1,500 sq. ft.) have a median price of \$270,000, while those without paved driveways average \$150,000.
- **Numerical Insight:** A paved driveway adds a median premium of \$120,000 for larger properties but shows minimal impact for smaller homes.
- **Insights:** Price variability for large properties with paved driveways is lower, suggesting a consistent added value in this segment.

#### 65. Seasonal Influence on Prices (Analysis of MoSold Across Years)

- **Results:** Properties sold in summer months (June and July) across multiple years show consistent price premiums of about 15%, with median prices around \$240,000. Winter sales average \$160,000.
- **Numerical Insight:** Seasonal price variation is consistent over years, showing a summer premium that raises median prices by \$80,000 compared to winter months..
- **Insights:** The highest variability is in July sales, with a standard deviation of \$45,000, reflecting increased market activity during summer.

#### 66. Remodeling Patterns and Price Impact

- **Results:** Properties remodeled within the last 10 years show a 20% price increase, with a median around \$270,000. Properties without remodeling average \$170,000.
- **Numerical Insight:** Recent remodeling yields a premium of approximately \$100,000, confirming the added value of updated properties.
- **Insights:** Price variance is high among recently remodeled properties, reflecting variability in other accompanying amenities.

#### 67. Pool Area Impact Beyond Pool Quality (PoolArea vs. PoolQC)

- **Results:** Properties with pool areas over 400 sq. ft. have a median price of \$320,000, while those with smaller pools average \$230,000, even when pool quality is consistent.
- **Numerical Insight:** Larger pool areas add an average premium of \$90,000 over smaller pools within the same quality rating.
- **Insights:** Price variability for larger pool areas is wide, suggesting that pool size combined with other features greatly influences high-value properties.

#### 68. Combined Analysis of Utility Access (Utilities) and Overall Property Size (GrLivArea)

- **Results:** Properties with full utilities and larger sizes (>2,000 sq. ft.) have a median price of \$300,000, while those with limited utilities and smaller sizes average around \$150,000.
- **Numerical Insight:** Full utility access increases the median price by approximately \$100,000 for larger homes, while smaller homes show minimal price impact.
- **Insights:** The price variance for larger properties with full utilities is narrower, indicating a consistent valuation for properties with comprehensive utility access.

## 69. Structural Consistency (HouseStyle vs. BldgType)

- **Results:** Single-family homes (1Fam) with a 2Story style have the highest median prices around \$240,000, while mixed-style buildings average around \$150,000.
- **Numerical Insight:** Properties with consistent structure types have a median premium of \$90,000 compared to mixed-style buildings.
- **Insights:** The standard deviation in prices for single-family 2Story houses is lower than for other styles, indicating a stable valuation for this popular structure.

## 70. Dynamic Clustering Using KMeans Based on Remodeling and Lot Features

- **Results:** KMeans clustering reveals three clusters: high-value properties (average \$310,000), mid-value (\$200,000), and lower-value properties (\$140,000), segmented by lot features and remodeling recency.
- **Numerical Insight:** High-value clusters are characterized by larger lot sizes and recent remodels, showing a \$170,000 premium over smaller or older properties.
- **Insights:** The variance within each cluster is lower than in the overall data, supporting clustering as a useful tool for grouping similar properties and enhancing model accuracy.

# Time Series

## 1. SARIMA Model

- **Trend Analysis:** SARIMA forecasts are generally lower than observed values, showing an underestimating trend, particularly evident in April 2010, where SARIMA forecasts 139,832, while the actual value is 171,344.
- **Peaks and Troughs:** The model projects a slight dip in prices by October 2009 (131,629), which aligns with a minor decrease in the actual trend but underestimates the true value of 171,022.
- **Deviation Timing:** Significant deviations appear in June 2010, with a forecast of 144,562 against an actual value of 171,458, highlighting a period where SARIMA did not capture an upward trend.
- **Accuracy Insights:** MAE of 29,151 indicates SARIMA's limited ability to track month-by-month variations closely, especially in high-variance months.

## 2. ETS (Exponential Smoothing) Model

- **Trend Analysis:** ETS forecasts closely align with observed values, with only minor overestimations and underestimations, reflecting good adaptability to both trend and seasonality.

- **Specific Month Accuracy:** For March 2010, ETS forecasts 161,542, while the actual value is 170,797, showing a slight underestimation. However, in November 2009, ETS overshoots with a forecast of 192,085 against an actual of 156,382.
- **Seasonality Capture:** The model captures seasonal patterns well, especially in December 2009 and June 2010, where forecasts are within 10,000 of actual values.
- **Error Distribution:** Consistent accuracy is seen across months, with an MAE of 17,548, reflecting reliable performance throughout the year.

### 3. Seasonal Naive Model

- **Trend Assumption:** Seasonal Naive assumes perfect repeatability of the previous year's pattern, projecting identical values for each month as observed last year.
- **Perfect Match:** For each forecasted month, such as August 2009 (165,671) or March 2010 (170,797), the forecast exactly matches the observed value of the previous cycle.
- **Trend Limitations:** No adjustments for trend or variability, so this model does not respond to any increases or decreases that may emerge within the forecasted year.
- **Error-Free Baseline:** As a benchmark, this model has zero MAE and MAPE because it does not deviate from last year's values.

### 4. Theta Model

- **Trend Analysis:** Theta demonstrates excellent tracking of observed trends, showing close alignment even in volatile months. For example, in October 2009, it forecasts 169,216 against the actual 171,022.
- **Accuracy in Peaks and Troughs:** Theta captures the fall in prices by July 2010, with a forecast of 168,074 against an actual value of 121,750, undershooting the decline but still following the trend.
- **Seasonality Responsiveness:** The model captures general seasonal trends, as seen in June 2010, where it forecasts 168,201 against an actual of 171,458, demonstrating minimal deviation.
- **Error Metrics:** With an MAE of 10,303 and MAPE of 6.83%, Theta provides one of the most accurate forecasts.

### 5. Simple Moving Average (SMA) Model

- **Flat Trend:** SMA provides a flat forecast based on the average of the previous 12 months (166,948), which does not account for specific monthly trends or seasonal changes.
- **Accuracy Impact:** This lack of adaptability results in higher errors in dynamic months; for instance, in April 2010, SMA forecasts a flat 166,948 against an actual value of 171,344.
- **Good Baseline Performance:** Despite its simplicity, SMA yields relatively low errors, with an MAE of 10,512 and a MAPE of 6.89%, making it a solid baseline.
- **Seasonality Limitations:** It does not capture peaks or troughs, as seen in July 2010, where it forecasts 166,948 against a much lower observed value of 121,750.

### 6. Bayesian Structural Time Series (BSTS) Model

- **Trend Responsiveness:** BSTS adapts to recent data trends, with forecasts smoothly following the trend. For instance, in March 2010, it forecasts 160,657, close to the observed 170,797.

- **Seasonal Adaptation:** The model captures seasonality moderately well, as seen in June 2010, forecasting 160,354 against the observed 171,458.
- **Error Distribution:** MAE of 13,242 shows it maintains moderate accuracy, though it lags slightly behind Theta and SMA.
- **Deviation in Downward Trends:** In July 2010, the model forecasts 159,969, slightly above the actual 121,750, suggesting some lag in capturing sharp declines.

Month	SARIMA	ETS	Seasonal Naive	Theta	SMA	BSTS
2010-08-01	135,004	179,634	165,671	169,470	166,948	160,359
2010-09-01	143,264	175,333	196,849	169,343	166,948	160,419
2010-10-01	131,629	171,661	171,022	169,216	166,948	160,132
2010-11-01	132,447	192,085	156,382	169,089	166,948	160,413
2010-12-01	133,587	181,074	164,015	168,962	166,948	160,328
2011-01-01	136,537	148,102	163,853	168,835	166,948	160,595
2011-02-01	148,782	162,984	174,823	168,709	166,948	165,258
2011-03-01	145,390	161,542	170,797	168,582	166,948	160,657
2011-04-01	139,832	150,438	171,344	168,455	166,948	160,196
2011-05-01	143,828	158,485	175,410	168,328	166,948	160,285
2011-06-01	144,562	164,755	171,458	168,201	166,948	160,354

2011-07-01	124,799	162,069	121,750	168,074	166,948	159,969
------------	---------	---------	---------	---------	---------	---------

## 1. Seasonal Patterns and Model Responsiveness

- **SARIMA and ETS:** Both SARIMA and ETS models capture seasonality, though with different sensitivities. ETS, for example, peaks in November 2010 with a forecast of 192,085, aligning with the trend of higher sales in late fall seen in the observed data. SARIMA, while seasonal, appears less responsive, particularly in February 2011 where it underestimates at 148,782.
- **Theta and SMA:** Both models provide stable, minimal variation forecasts, suggesting these methods may be better suited for data with less pronounced monthly fluctuations or for long-term average predictions.
- **Seasonal Naive:** This model perfectly replicates last year's cycle, highlighting any exact repeat of past seasonal cycles but lacking any adaptability for trend changes.

## 2. Trend Observations Across Models

- **SARIMA and BSTS:** Both models hint at a slight downward trend overall, forecasting lower average prices in 2011 compared to 2010. For instance, SARIMA's July 2011 forecast is 124,799, a significant drop from prior months, which aligns with the observed July 2010 dip.
- **ETS and Theta:** These models maintain a more stable forecast, implying either a plateau in prices or only minor variations in average monthly values. Theta, for example, stays between 168,074 and 169,470 across all months.

## 3. Forecast Uncertainty and Confidence

- **ETS, SARIMA, and BSTS:** These models adjust to short-term price trends, making them useful when expecting gradual market shifts. ETS is more adaptive with seasonality, while BSTS provides a balanced approach by slightly adjusting monthly values.
- **SMA and Seasonal Naive:** These provide simple benchmarks; however, they don't respond to changes in trends or seasonality, offering limited value if unexpected market shifts occur.

## 4. Potential Decision-Making Applications

- **Short-Term Forecasts:** For decision-making over the coming year, models like **ETS and BSTS** could provide more actionable insights by capturing both monthly seasonality and potential trend changes.
- **Long-Term Forecasts:** **Theta and SMA** models might be more suitable if the focus is on understanding the average price level rather than monthly fluctuations, as they offer a stable prediction reflecting overall market tendencies.

# Predictive Modelling(Advanced Regression)

## 1. Model-Specific RMSE Evaluation

- **ElasticNet:** Validation RMSE = 0.4004
  - **Insight:** ElasticNet, while useful for handling multicollinearity and feature selection through regularization, shows the highest RMSE among all models. This indicates that a linear model with regularization does not fully capture the complex non-linear relationships in the dataset, resulting in a lower predictive accuracy.
- **RandomForest:** Validation RMSE = 0.1544
  - **Insight:** RandomForest achieves a low RMSE, suggesting it effectively models complex relationships and interactions within the data. However, as a non-boosted ensemble method, it may still lack the refined performance that boosting algorithms provide, particularly on more nuanced patterns.
- **GradientBoosting:** Validation RMSE = 0.1590
  - **Insight:** Gradient Boosting performs slightly less accurately than RandomForest but provides robust results. This method incrementally improves predictions through a sequential boosting approach, which likely captures non-linear interactions effectively. The close RMSE to RandomForest implies that it captures similar features but with different emphasis due to its boosting nature.
- **XGBoost:** Validation RMSE = 0.1600
  - **Insight:** XGBoost, another boosting algorithm, achieves an RMSE similar to GradientBoosting. Known for its regularization capabilities and scalability, XGBoost performs well but slightly underperforms compared to GradientBoosting and RandomForest in this dataset. This minor difference may be due to parameter tuning limits or dataset-specific intricacies that XGBoost didn't fully capture.

## 2. Optimized Blended Model Analysis

- **Blended Model (RandomForest + GradientBoosting):** Validation RMSE = 0.1529
  - **Optimal Weights:** RandomForest = 0.3163, GradientBoosting = 0.6837
  - **Insight:** The optimized blended model, which combines RandomForest and GradientBoosting, achieves the lowest RMSE (0.1529). This improvement indicates that a weighted blend of these two models successfully leverages the strengths of both:
  - **RandomForest** contributes robustness and stability in capturing larger patterns and interactions.
  - **GradientBoosting** provides incremental refinements, enhancing the model's ability to capture subtle variations in the data.
  - **Weight Analysis:** With a higher weight for GradientBoosting (68.37%) than RandomForest (31.63%), the blend relies more heavily on the boosting model's iterative improvements while benefiting from the random forest's ensemble.