

# Navigating the Real Estate Maze: The Confluence of Zoning, Condition, Size, and Historical Trends on Property Values

## Data Abstract Description:

The dataset in question is a structured collection of data where each row represents a single real estate transaction, detailing various attributes of the property sold. The columns, or variables, provide extensive information about each property's features, conditions, and sale details.

## Dataset Overview:

Subjects (Rows): Each row corresponds to an individual property involved in a sale.

Variables (Columns): The dataset comprises numerous columns, each representing specific characteristics or features of the property.

## Physical Characteristics:

The housing dataset captures a wide array of physical characteristics of each property. This includes the building class, which describes the type of dwelling, and the architectural style of the house. The dataset also includes details on the construction materials used, such as the type of roof and foundation. The age of the property is captured through the original construction date and any remodel dates. Interior features are also well-documented, including the basement condition and finished areas, the type and quality of the heating system, the presence of central air conditioning, and the electrical system. The size of the property is detailed through the square footage of each floor, the number of bathrooms and bedrooms, and the total number of rooms. The quality of key interior components, such as the kitchen and any fireplaces, is also rated.

Finally, the dataset includes information on the presence and size of garages, decks, porches, and pools.

### Location and Accessibility:

The dataset provides key insights into the location and accessibility of each property. This includes the zoning classification, which indicates whether the property is in a residential, commercial, or other type of area. The specific neighborhood is also noted. The dataset captures the type of road access, such as whether the property is on a public street or has access via an alley. The proximity to various amenities or conditions, such as being near a park or a railway, is also included. These locational attributes can play a significant role in determining the desirability and value of a property.

### Utilities and Amenities:

The housing dataset details the availability and type of utilities for each property, such as whether it has access to public electricity, gas, and water supply. The presence of amenities that can enhance the value and livability of a property is also captured. This includes whether the property has a garage and the size and quality of the garage. The dataset notes the presence of a pool and rates the pool's quality. Other amenities like the presence of a porch, deck, or air conditioning are also included. These features can significantly impact the appeal and value of a property to potential buyers.

### Interior Features and Quality Rating:

The interior of each property is described in detail in the housing dataset. This includes the number of rooms and the overall room configuration, such as the number of bedrooms and bathrooms. The dataset also provides quality ratings for key interior components. The kitchen, for example, is rated for its overall quality. If the property has any fireplaces, their quality is also assessed. The condition and quality of the heating system are also noted. The dataset also includes an overall assessment of the property's condition and functionality. These interior features and quality ratings provide insight into the living conditions and potential renovation needs of each property.

### Sale Details:

The housing dataset includes key information about the sale of each property. Most importantly, it includes the sale price, which is often the target variable in housing price prediction tasks. The

dataset also notes the type of sale, such as whether it was a new construction, a foreclosure, or a short sale. The specific month and year of the sale are also included. Any unique conditions of the sale that could affect the price, such as whether the sale was made under duress, are also noted. These sale details provide essential context for understanding the pricing and market conditions for each property transaction.

## Data Types:

**Numerical:** Variables that are quantitative and measurable, such as LotArea, YearBuilt, TotalBsmtSF, 1stFlrSF, and SalePrice.

**Categorical:** Variables that are qualitative and describe property characteristics, such as MSZoning, Neighborhood, HouseStyle, and SaleCondition.

**Ordinal:** Variables that represent ratings or rankings, such as OverallQual, ExterQual, and BsmtCond, where the numbers correspond to different levels of quality or condition.

1. Physical Characteristics: MSSubClass, OverallQual, YearBuilt
2. Location and Accessibility: MSZoning, Neighborhood
3. Utilities and Amenities: HeatingQC, CentralAir, PoolArea, PoolQC
4. Interior Features and Quality Ratings: BsmtFinSF1, BsmtQual, KitchenQual
5. Sale Details: SalePrice, SaleType, SaleCondition

## Used Variables

**Building Quality (OverallQual):** Rates the overall material and finish of the house on a scale from 1 to 10.

**YrSold (YearBuilt):** Indicates the year the house was sold.

Zoning Classification (MSZoning): Specifies the general zoning status of the sale (like residential, commercial).

Sale Price (SalePrice): The transaction price, which is typically the dependent variable in price prediction models.

GrLivArea : Above grade (ground) living area square feet.

## Possible Statistical Use Cases:

Regression Analysis: To predict the SalePrice based on various predictors like OverallQual, YearBuilt, and GrLivArea.

Classification: To categorize homes into different tiers or predict market segments based on features like Neighborhood and OverallCond.

Time-Series Analysis: Examining trends over time, especially with variables like YrSold and MoSold, to analyze seasonal effects on market dynamics.

## Question 1 :The Zoning Effect

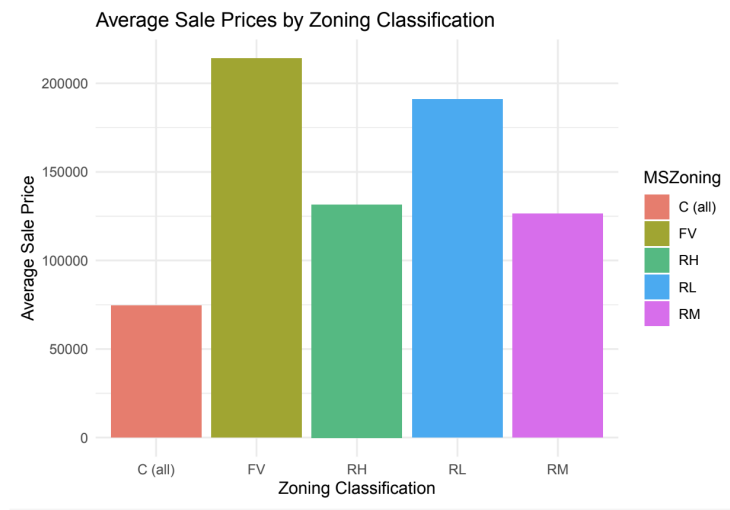
Research Question-Impact of Zoning on Property Values: How do different zoning classifications (e.g., Residential, Commercial, Agriculture) impact the sale prices of properties?

The journey begins with an exploration of how zoning regulations shape the real estate landscape. Zoning laws, which designate specific areas for residential, commercial, or agricultural use, play a pivotal role in determining property utility and potential. By analyzing sales data across various zones, the study highlights significant disparities in property values. Residential zones featuring low-density housing often attract higher property values due to their appeal to families looking for more spacious living environments, while commercial zones showcase a premium on accessibility and proximity to economic activities.

Predictor Variables: MSZoning (zoning classification)

Target Variable: SalePrice (property sale price)

## Bar Graph



## Interpretation

We start by investigating the impacts of what zoning classification the houses that were sold fall under. This lets us start with a fairly broad perspective that doesn't require as much knowledge of the specific neighborhood/geographic location where these houses were sold (Ames, Iowa). Zoning laws and classifications have a substantial impact on the development of towns, and cities, but how is that reflected, if at all, in the prices of homes themselves? Note the following zoning categories: C(all) encompasses all types of Commercial zones, FV is Floating Village Residential, and RH, RM, and RL, are Residential

High, Medium, and Low Density zones. Homes in the Floating Village Residential Zone are different from typical homes in that they're not built on land, but are floating on a body of water.

The above bar graph compares the average sale prices across different zoning classifications. It seems the highest average house sale prices tend to be in Floating Village Residential zones and Residential Low Density zones. Meanwhile, Residential High and Medium Density zone homes were both lower relative to Floating Village Residential and Residential Low-Density zone homes, but still sold for much more than any homes in commercial zones.

In conclusion this illustrates the average sale prices across different zoning classifications. Residential zoning (FV, RL, and RM) shows higher average sale prices, with Floating Village Residential (FV) being the highest, suggesting a premium for such properties. Commercial zoning (C) has the lowest average price, indicating a possible preference for residential spaces or less value in commercial areas in the context of this dataset.

## Statistical Test / Analyses for Bar Graph

A chi-square test is used to determine if there is a significant association between two categorical variables.

Null Hypothesis ( $H_0$ ): The number of houses is uniformly distributed across the different zoning classifications. This means that each zoning classification has the same number of houses, indicating no difference in frequency among the categories.

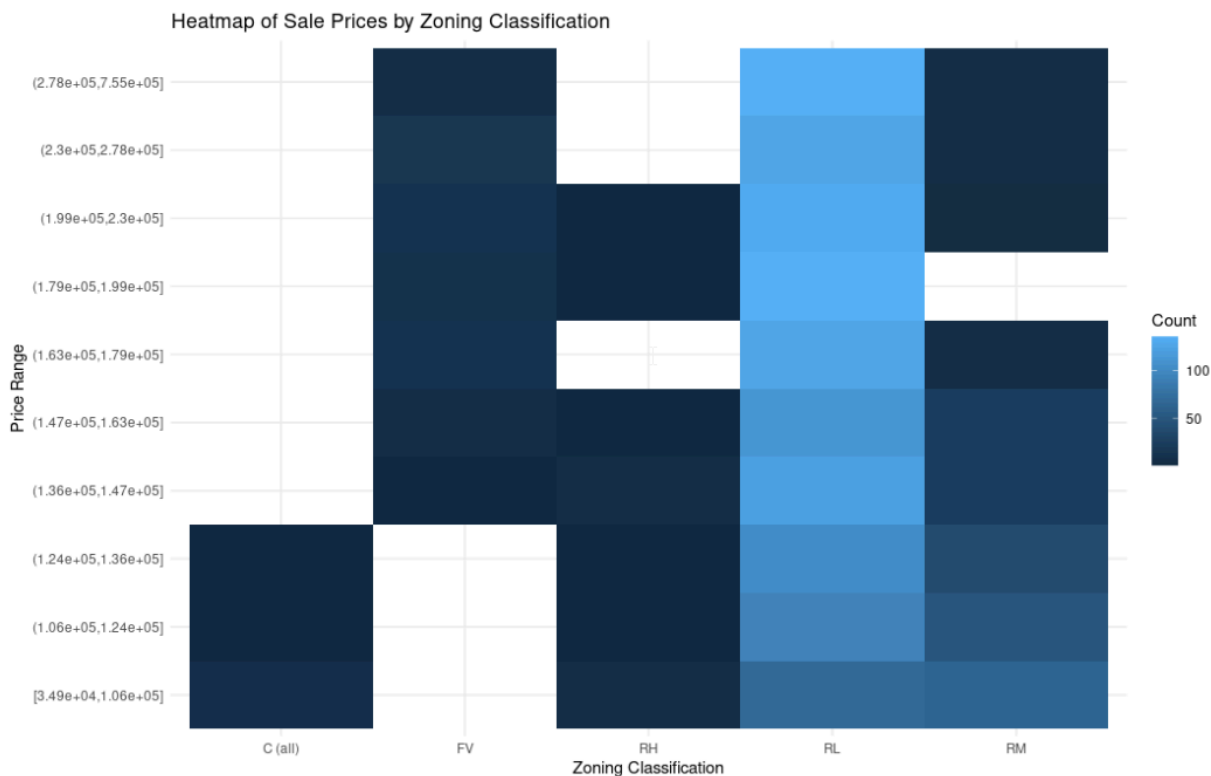
Alternative Hypothesis ( $H_1$ ): The number of houses is not uniformly distributed across the different zoning classifications. This implies that some zoning classifications have more or fewer houses than others, indicating a difference in frequencies among the categories.

### Chi-squared test for given probabilities

```
data: zoning_counts$n  
X-squared = 3255.4, df = 4, p-value < 2.2e-16
```

The chi-squared test displayed above shows that the test statistic is calculated as 3255.4. This value indicates the difference between observed and expected frequencies. The degrees of freedom is 4 which is determined by  $n-1$ . The p-value associated with the test has extremely small statistical significance. Hence, the test shows strong evidence to reject the null hypothesis. As the P value is less than the commonly accepted alpha of 0.5 we reject the null.

## Heatmap Graph



## Interpretation

The heatmap above helps to better understand how the sale prices of these homes vary by zoning classification and also shows the case of the low sample size for some of these zoning categories. The

lighter shades represent a higher count of homes sold at a particular price range. Based on the heat map above we can see that the vast majority of homes fall in the Residential Low Density zone. Interestingly, we can also see that there is no house in the commercial zone that sold for over 150,000 dollars and there is no house in Floating Village Residential zones that sold under 150,000. We can conclude from this that most typical homes sold in Ames, Iowa are in low-density residential zones and are expensive relative to other zones. Overall, the heatmap suggests that residential zones (RL and RM) have more sales and a wider distribution of sale prices, while FV, as a more specialized zone, has fewer but higher-priced sales. The commercial zone (C) has the least activity with lower sale prices.

## Question 2: The Condition and Size Paradigm

Research Question : Effects of Property Condition and Size on Sale Prices-How does the condition of the property, as well as its size (e.g., lot area, number of rooms), affect the sale price?

Transitioning from the macro influences of zoning, the narrative shifts focus to the micro-level details of individual properties. This chapter examines how the physical condition and size of properties influence their market value. Through statistical analysis, it is observed that well-maintained properties with updated features fetch higher prices, underscoring the importance of condition. Simultaneously, the size of the property, both in terms of lot area and livable space, is found to have a proportional relationship with sale prices. Larger properties command higher prices, but the increment varies significantly across different neighborhoods

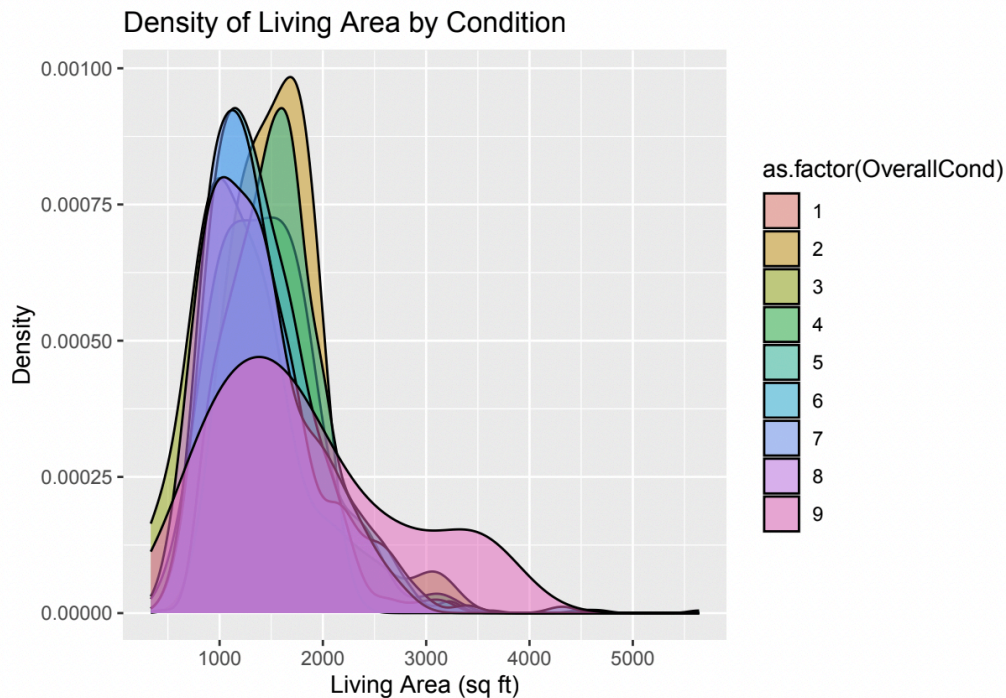
Effects of Property Condition and Size on Sale Prices:

Predictor Variables: `OverallCond` (overall condition of the house), `GrLivArea` (above grade living area square feet)

Target Variable: `SalePrice`

## Density Plot



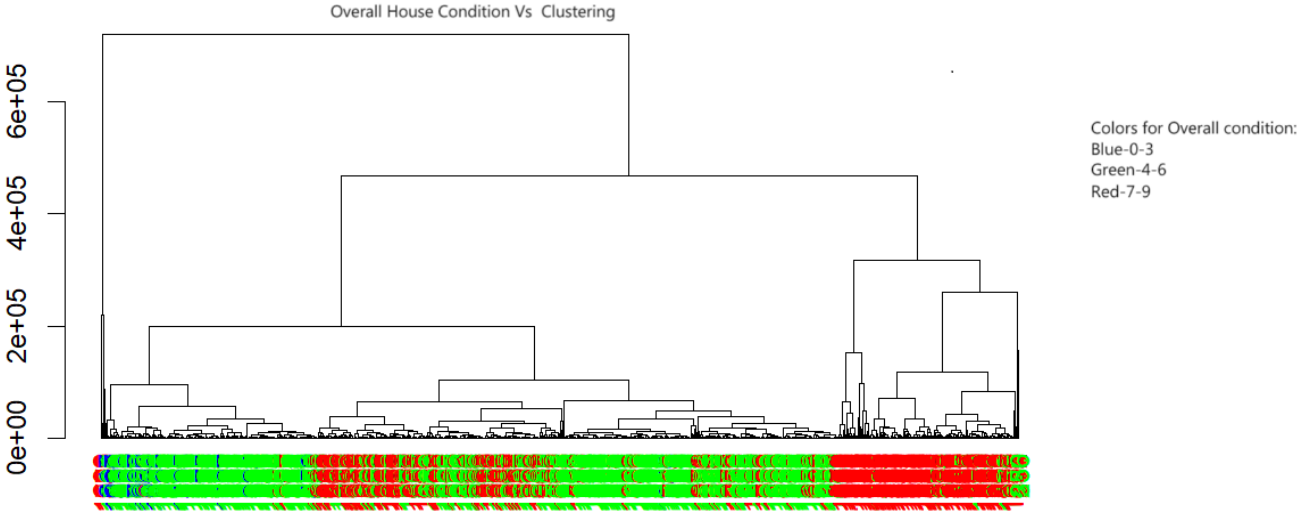


## Interpretation

Based on the density plot provided, there appears to be a clear relationship between a property's living area and its overall condition rating. The distribution shifts rightward as the condition improves from 1 to 9, indicating that homes in better condition tend to have larger living areas on average. Properties rated as a 1 (presumably the worst condition) are concentrated at the lowest end of the living area range, mostly under 1000 sq ft. In contrast, those with the highest condition ratings of 8 or 9 are skewed towards the larger end, peaking around 3000 sq ft and extending past 5000 sq ft. The intermediate rankings from 2 to 7 show a steady progression, with each successively better condition category having a higher proportion of larger homes. This positive association between condition and size likely reflects a combination of factors - newer homes being larger and in better shape, owners of bigger properties having more resources to maintain them well, and renovations/expansions that simultaneously add living space and improve condition. From an investment perspective, this suggests that properties with both sizable living areas and strong condition ratings will command the highest prices, while smaller, lower condition homes offer more affordable entry points but may require additional improvement costs. Further analysis correlating sale prices to these attributes would help quantify their impact on property values. Overall, this graph indicates that the majority of properties are of average

condition with a broad range of living area sizes. In contrast, properties in very high or very low condition ratings tend to have smaller and less variable living areas.

## Dendrogram



## Interpretation

The dendrogram presents a hierarchical structure of property clusters based on a range of quantitative features, and is categorized by house condition. Properties are grouped by similarity, with those sharing more characteristics placed closer together. The color-coding reflects overall house conditions, with blue for lower conditions (1-3), green for moderate (4-6), and red for higher conditions (7-9).

Houses with higher conditions (red) form tight clusters, indicating a strong similarity among these properties, suggesting that excellent condition homes have consistent features that set them apart from others. Properties in the blue and green categories show greater dispersion, implying more variability within these groups in terms of the features considered.

Larger clusters, particularly those combining different condition ratings, highlight that despite varying conditions, there is a set of shared characteristics influencing the grouping. This means that factors other than the overall condition, potentially location, size, or age, play a significant role in grouping these properties.

## Question 3 :The Historical Value Paradigm

Research Questions-Historical Value Assessment: How have property values changed over the years, and what historical events or market conditions might have influenced these changes?

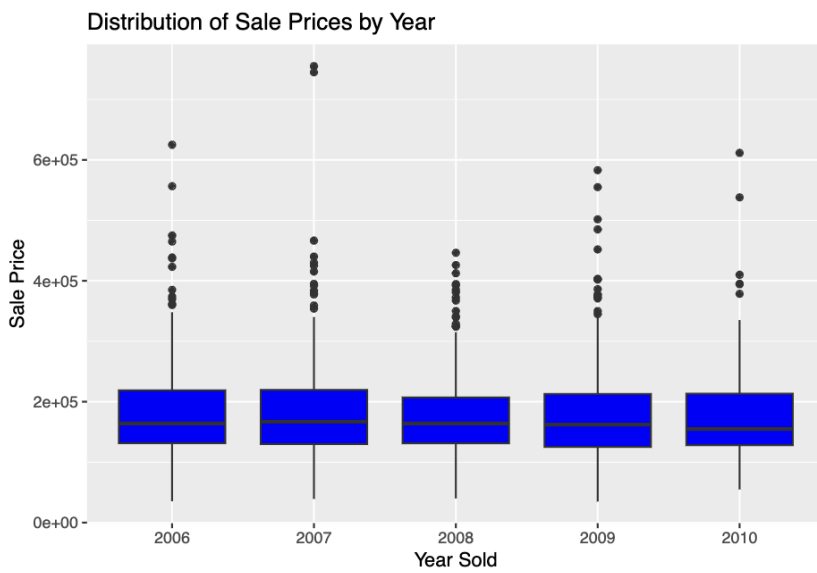
Moving beyond the intrinsic attributes of properties, this chapter delves into the temporal dynamics of the real estate market. Here, the analysis focuses on how historical trends and economic factors over time have shaped property valuations. The year in which a property is sold (YrSold) serves as a temporal marker to the flow of the market.

It's discovered through other research that certain years see spikes in value, possibly aligned with economic booms, while other periods show dips, potentially due to recessions or market corrections. Additionally, the influence of historically significant events, such as policy changes or infrastructure developments, is observed to have lasting effects on market trends. The interplay of `YrSold` with variables that represent economic conditions—such as interest rates, employment data, or stock market indices—provides a nuanced understanding of the market's responsiveness to broader economic climates.

Predictor Variables: YrSold (year sold)

Target Variable: SalePrice

Box Plot

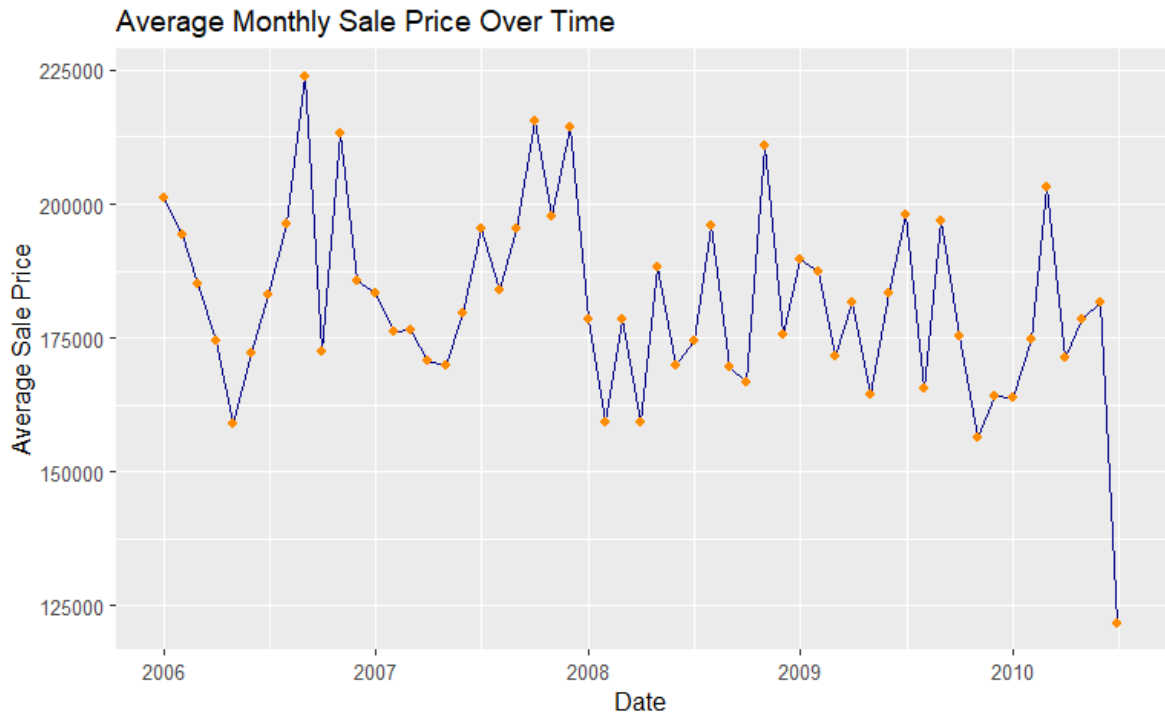


1

## Interpretation

The boxplot displays the sale price distributions of properties from 2006 to 2010, revealing a stable real estate market over these years. The median sale prices, indicated by the horizontal lines within the boxes, show little year-to-year variation, suggesting consistent market conditions. The interquartile ranges are also similar across the years, pointing to a steady spread in the middle 50% of the data, with no significant shifts in market volatility. Outliers are present for each year, with 2007 and 2008 exhibiting some exceptionally high-priced sales, yet these do not appear to disrupt the overall stability of the market. The consistent range of prices, as demonstrated by the lengths of the 'whiskers', further underscores a market without drastic fluctuations in property sale prices during this five-year span.

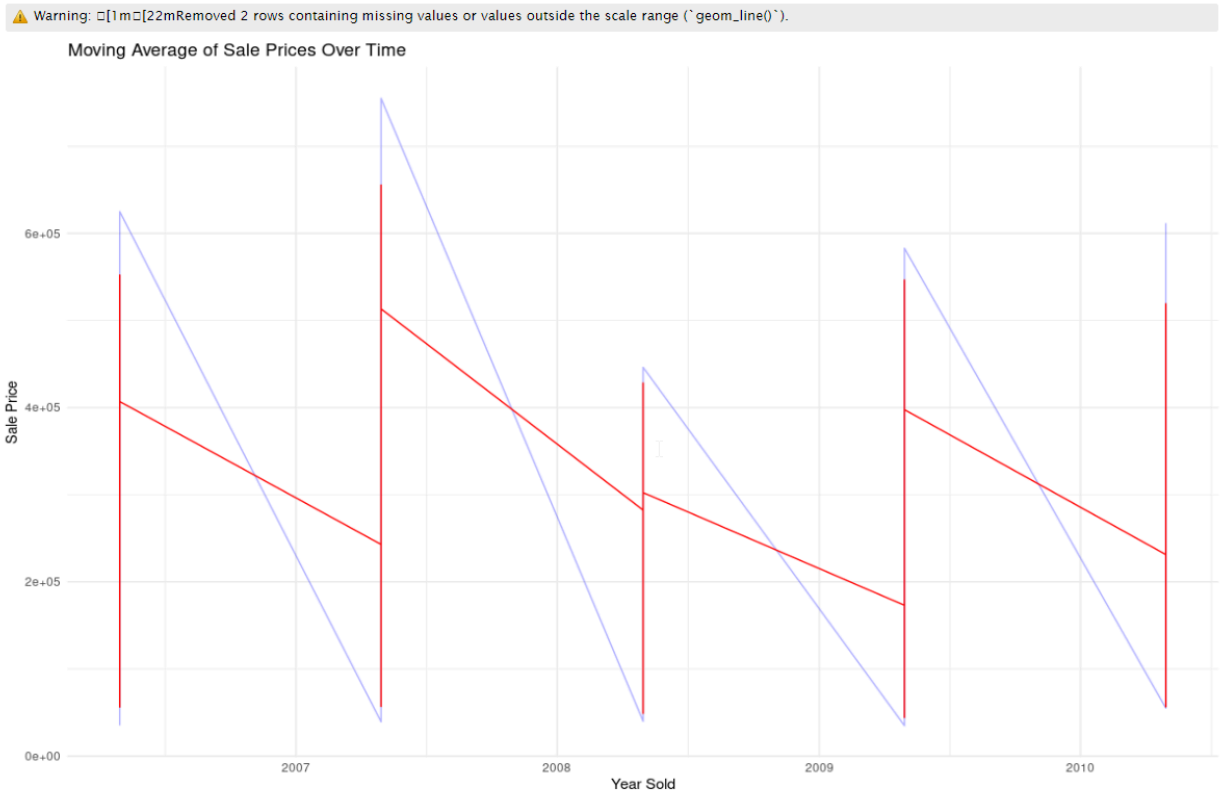
## Time Series



## Interpretation

The Time graph depicts fluctuations in average monthly sale prices of properties over a span from 2005 to 2010, revealing a pattern indicative of seasonality with regular peaks and troughs each year. This suggests that the property market experiences predictable cycles of activity, with certain months consistently yielding higher or lower sale prices. Although the general trend of sale prices remains relatively stable over the years, a period of heightened volatility is evident around 2007 and 2008, where the peaks are more pronounced. This could be reflective of economic factors influencing the market at that time. Notably, there's a significant dip in average sale prices toward the end of 2010, signaling either a market downturn or a potential outlier in the data. The graph captures the dynamic nature of the housing market, subject to seasonal influences and broader economic conditions. The repeating patterns of rise and fall in the average sale prices could be linked to the traditional cycles in real estate activity, with certain months perhaps seeing more sales activity due to seasonal buying and selling behaviors.

# Moving Averages Graph



## Interpretation:

The graph displays the moving average of sale prices (red line) overlaid with actual sale prices (blue line) from 2006 to 2010. The moving average smooths out the volatility seen in the actual sale prices, revealing underlying trends in the real estate market. Notably, there is a significant peak in prices around 2008, followed by a sharp decline, which likely correlates with the global financial crisis. By 2010, the moving average began to stabilize, indicating signs of market recovery or stabilization after the downturn.

## Conclusion:

The exploration of the housing dataset offers substantial insights into the factors that impact property values across various dimensions—zoning, physical attributes, and temporal market dynamics. Through the analysis:

**Zoning's Impact:** We discovered that zoning classifications significantly influence sale prices, with residential zones generally fetching higher prices than commercial ones. The rarity and uniqueness of Floating Village Residential zones particularly command premium prices.

**Property Condition and Size:** The data reveals a clear positive correlation between property size, condition, and sale prices. Well-maintained and larger properties tend to attract higher prices. This relationship underscores the importance of property condition in real estate valuation and highlights the market preference for spacious and well-kept homes.

**Temporal Dynamics:** The analysis of yearly and monthly sale trends illustrates the stability of the real estate market over time, despite minor fluctuations due to economic factors. Seasonality appears as a consistent factor, influencing buying patterns and property values throughout the year.

## Future Research Questions

1. **Long-term Economic Impacts:** Regarding further influence of zoning effect, we are interested in how broader economic conditions, such as changes in employment rates, inflation, and interest rates, impact housing prices over extended periods? Further research could integrate economic data with housing market trends to develop more comprehensive predictive models.
2. **Urban vs. Suburban Trends:** How do property values and features differ between urban and suburban settings? Future research could examine the shifting

dynamics between these areas, especially in light of changing work patterns post-pandemic. Influenced by Question 1

3. Demographic Shifts and Economic Impacts: How do changes in demographics, such as aging populations or varying family sizes, influence housing needs and market trends? Analyzing these shifts could provide insights into future demand for different types of housing. Influenced by Question 2
4. Housing Market Resilience: Following economic shocks or natural disasters, how resilient are housing markets in different regions? Studies could examine recovery times and factors that contribute to quicker recoveries in housing markets post-crisis. Influenced by Question 3
5. Affordable Housing Policies: What are the long-term effects of different affordable housing policies on market prices and housing accessibility? Future research could assess the effectiveness of subsidies, rent controls, and incentive programs for developers to include affordable units in new projects. Influenced by Question 3