Proceedings of CONF-SPML 2027 Symposium: Intelligent Network Security and Machine Learning for Communication Systems
DOI: 10.54254/2755-2721/2027.MELB37705

A Data-Driven Analysis of Housing Consumption Patterns: Evidence from King County

Zihan Huang

Guangzhou Tieyi Middle School, Guangzhou, China
Huangzihan27@163.com

Abstract. Housing is one of the biggest expenses for most households, so changes in how people consume housing can tell us a lot about broader trends in consumption and upgrading. This paper uses a dataset of housing sales in King County, Washington, to explore what factors influence housing prices and what these patterns say about consumer behaviour. The analysis uses descriptive statistics, correlation analysis, and multiple linear regression to identify the key drivers of housing prices. The results show that living area ($r = 0.702$) and waterfront status are the strongest predictors of price, followed by property grade and view quality. However, the year the house was built does not have a strong effect on price once other factors are accounted for, which suggests that consumers care more about quality and location than about whether a property is brand new. These findings suggest that housing consumption upgrading is less about buying newer houses and more about buying better houses, which has more space, better views, and higher quality construction.

Keywords: housing prices, consumption upgrading, hedonic pricing model, King County, data analysis

1. Introduction

Recent research has increasingly utilized large-scale housing transaction datasets to explore price determinants and their reflection of consumer behavior [1-3]. The hedonic pricing model has become a standard framework for decomposing housing prices into structural, locational, and neighborhood attributes, treating a house as a bundle of features whose value is reflected in its price [4]. Previous studies have shown that housing prices are closely linked to consumption upgrading [1-3]. J. P. Franklin et al. applied this model to King County data and found that accessibility to employment affects property values [5]. As the largest household expenditure, housing choices reflect deeper shifts in consumer priorities and provide a useful window into broader consumption upgrading trends. The use of transaction data enables more robust empirical analysis than traditional small-sample surveys.

This study analyzes the real estate transactions in King County, Washington, to identify the factors influencing housing prices and the underlying consumer preferences they reveal. This study contributes to the broader discussion on consumption upgrading and offers practical implications for planners and developers.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2027.MELB37705

2. Data description

2.1. Data source and sample

This study uses the King County housing sales dataset, which contains a number of observations of residential property transactions in King County, Washington, including the city of Seattle. The dataset includes information about both structural features (like living area, number of bedrooms, and number of bathrooms) and locational features (like zip code, latitude, and longitude). The data comes from publicly available real estate transaction records and has been used in previous studies on housing economics [6, 7].

2.2. Variables

The dependent variable is price, which is the final sale price of each property in US dollars.

The independent variables include structural characteristics, such as the square footage of the interior living space, the number of bedrooms and bathrooms, the number of floors, and the year the house was built. Notablly, grade and condition are complementary measures of house quality. These variables were chosen based on previous research and economic theory [6, 7]. Grade captures the intrinsic construction quality and architectural design, while condition reflects the current state of maintenance. Both are expected to positively influence price, as consumers generally prefer well-built and well-maintained properties. sqft_living measures the size of the property, which is expected to have a positive relationship with price. Waterfront status is expected to have a large positive effect because such properties are scarce and water views are highly desirable. View quality and grade reflect the aesthetic and construction quality of the property, which are also expected to have a positive influence on price. Condition captures the current state of maintenance, and yr_built captures the age of the property.

2.3. Data preprocessing

A quick check of the data showed no missing values for the main variables used in this analysis. However, there were some extreme outliers in price and sqft_living. These were kept in the analysis because they represent real high-end properties rather than data errors. Previous studies using the same dataset have taken a similar approach, noting that luxury properties provide useful information about the upper end of the housing market [6, 7].

Specifically, properties with prices above the 99th percentile or below the 1st percentile were examined individually. No properties were found to have data entry errors, such as implausibly low prices for large homes or implausibly high prices for small homes. For living area, properties with square footage exceeding 5,000 were also retained, as these represent legitimate luxury properties rather than measurement errors. This conservative approach to outlier treatment ensures that the regression results reflect the full spectrum of the housing market.

2.4. Descriptive statistics

The average housing price in King County is about $540,000, with a wide range from under $100,000 to over $3,000,000. The large standard deviation of $340,000 shows that there is a lot of variation in housing prices across the county. The average living area is 2,080 square feet, and a typical property has 3 bedrooms and 2 bathrooms. The average grade is around 7, which means a standard level of construction quality. The average condition rating of 3 suggests that most

DOI: 10.54254/2755-2721/2027.MELB37705

properties are reasonably well-maintained. Only 0.75% of properties are waterfront, highlighting their scarcity and helping explain the large price premium observed in previous research [6]. The average year built of approximately 1969 indicates a relatively aged housing stock.

3. Results and analysis

3.1. Price distribution

The housing price distribution is strongly right-skewed. Most properties are in the lower-to-middle price range, with a long tail of high-value properties exceeding $1,000,000. The median price of $450,000 is significantly lower than the mean of $540,000, confirming that a small number of expensive properties pull the average upward, a common feature of housing markets [6]. This polarization reflects a key aspect of consumption upgrading, where a segment of consumers invests substantially more in premium properties. Figure 1 shows the histogram of housing prices, which clearly illustrates the right-skewed distribution.

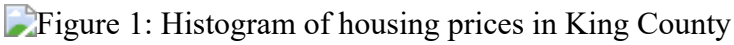

Figure 1. Histogram of housing prices in King County

3.2. Correlation analysis

The correlation between living area and housing price is $r = 0.702$ ($p < 0.001$), which indicates a strong positive relationship. This means that living area alone explains about 49% of the variation in housing prices. Among all the variables in the dataset, living area is the single most important predictor of housing value.

Figure 2 shows the scatterplot of living area versus housing price. The plot shows a clear positive relationship, though it is not perfectly linear. The variation in price gets larger at higher levels of living area, which suggests that for larger homes, other features start to matter more in determining the price.

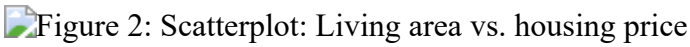

Figure 2. Scatterplot: living area vs. housing price

3.3. Waterfront premium

Waterfront properties sell for much higher prices than non-waterfront ones. The median price for waterfront properties is about $1,450,000, compared to $450,000 for non-waterfront properties — a difference of over 200%. Figure 3 shows a boxplot comparing waterfront and non-waterfront housing prices.

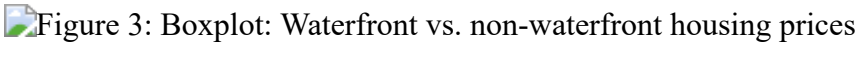

Figure 3. Boxplot: waterfront vs. non-waterfront housing prices

This result is consistent with Munaiah et al., who reported a waterfront coefficient of $554,704 [6]. Waterfront properties constitute less than 1% of all properties, and demand for water views is high. Research consistently finds that access to natural amenities such as coastlines and lakes is associated with higher property values [8]. This premium reflects a key aspect of consumption

DOI: 10.54254/2755-2721/2027.MELB37705

upgrading: consumers increasingly pay for environmental amenities and lifestyle benefits rather than just basic shelter.

3.4. Property condition and grade

Properties with higher grade ratings — which indicate better construction quality — are linked to significantly higher prices. The grade variable has a particularly strong relationship with price, which matches the finding in, that grade correlates with price at r = 0.667 [6]. This suggests that consumers are willing to pay more for better build quality and design.

3.5. Age and renovation

The relationship between property age and price is more complicated. Newer properties tend to sell for more, but the relationship is weak and not consistent. Many older properties, especially those in good locations, keep their value even after many decades. Chen found that certain structural variables (such as the no. of floors) may not significantly impact prices, while features such as living standards and waterfront remain the main driving factors. This suggests that consumers care more about whether a property has been updated than about how old it is. People are willing to buy older homes if they have been renovated.

3.6. Multiple linear regression analysis

To isolate the effect of each attribute, a multiple linear regression model was estimated:

$$price = \beta 0 + \beta 1 * sqft_living + \beta 2 * bedrooms + \beta 3 * bathrooms + \beta 4 * floors \quad (1)$$

$$+\beta 5 * waterfront + \beta 6 * view + \beta 7 * condition + \beta 8 * grade + \beta 9 * yr_built + \varepsilon$$

The regression results are shown in Table 2. The model explains about 64.5% of the variation in housing prices ($R^2 = 0.645$), which is close to the 65.3% [6].

Table 2. Multiple linear regression results

Variable	Coefficients	Standard Error	t Stat	P-value
sqft_living	161.6763	13.7990	11.7165	0.0000
bedrooms	-26142.4620	9128.0686	-2.8640	0.0043
bathrooms	34128.6844	14422.5580	2.3663	0.0182
floors	26225.8696	14176.9561	1.8499	0.0646
waterfront	672028.3865	75887.7537	8.8556	0.0000
view	64738.5483	9392.9777	6.8922	0.0000
condition	-3734.9571	9626.1614	-0.3880	0.6981
grade	105574.4536	9189.2780	11.4889	0.0000

DOI: 10.54254/2755-2721/2027.MELB37705

Table 2. (continued)

yr_built	-3635.1828	282.3288	-12.8757	0.0000

The signs and sizes of the coefficients are mostly what we would expect. The strongest predictor is living area: each additional square foot adds about $161.68 to the sale price. This effect is highly significant ($p < 0.001$) and matches the correlation results. Waterfront status is the second strongest predictor, adding about $672,028 to the property value.

The coefficient for bedrooms is negative: for a given living area, more bedrooms lowers the predicted price, possibly because it signals smaller rooms and lower overall quality. This counterintuitive finding suggests that consumers prioritize living space per room rather than the raw number of rooms. A property with fewer but larger bedrooms may be perceived as more luxurious and functional than one with many cramped bedrooms.

Floors is the only variable that is not statistically significant at conventional levels ($p = 0.0646$). This suggests that once other factors such as total living area and property grade are accounted for, the number of levels does not independently affect price. One possible explanation is that consumers are indifferent between single-level and multi-level homes, or that any preference for one over the other is already captured by other variables such as living area and view quality.

The negative coefficient for yr_built shows that older houses tend to sell for less, all else being equal, which aligns with economic intuition: newer homes incorporate modern building codes, materials, and design trends that appeal to buyers. However, the effect is modest: each additional year of age reduces the predicted price by only about $3,635, suggesting that age alone is a relatively minor factor compared to living area or location. This also implies that well-maintained older homes in desirable areas can effectively compete with newer construction.

To sum up, housing prices are influenced by supply constraints and consumer demand, emphasizing the importance of living area and waterfront status, as they provide significant utility. This aligns with consumer theory's revealed preference concept. Sachdeva et al. found that housing attribute effects vary across locations in King County, meaning the same features may have different values depending on neighborhood [9]. Gorjian reviewed hedonic regression and machine learning approaches, concluding that hedonic regression remains useful because it provides interpretable results for understanding consumer preferences [10].

4. Conclusion

This study analyzes King County housing transaction to identify key price determinants and their implications for consumption upgrading. Results show that prices are primarily driven by living area, waterfront status, grade, and view quality, while the year built has a relatively weak effect. The regression model explains 64.5% of price variation, confirming that structural and locational features are the main determinants of housing value.

These findings provide empirical evidence that housing upgrading is multidimensional—encompassing space, quality, and amenities, rather than a simple shift toward newer homes. The substantial premium associated with waterfront locations and higher-grade construction suggests that modern consumption increasingly involves lifestyle and environmental factors, offering practical guidance for urban planners, developers, and prospective buyers.

However, the data in this study excludes neighborhood characteristics such as school quality and crime rates. Future research could address these by adding location-specific variables, using spatial regression, or comparing across markets. Nevertheless, housing consumption upgrading is best

DOI: 10.54254/2755-2721/2027.MELB37705

understood as a move from "having a house" to "having a better house," reflecting broader shifts in consumer behavior.

References

[1] Ding, Y., Chin, L., & Taghizadeh-Hesary, F. (2024). Urban housing prices and consumption structure upgrading: Different relationships in urban and rural China. *International Journal of Strategic Property Management, 28*(5), 334–348.

[2] Wang, X. (2024). *The impact of housing prices on the upgrading of consumption* [Master's thesis]. University of Sydney.

[3] Zheng, H., Qian, J., Liu, G., Wu, Y., Delang, C. O., & He, H. (2023). Housing prices and household consumption: A threshold effect model analysis in central and western China. *Humanities and Social Sciences Communications, 10*(1), 764.

[4] Malpezzi, S. (2003). Hedonic pricing models: A selective and applied review. In *Housing economics and public policy* (pp. 1–39).

[5] Franklin, J. P., & Waddell, P. (2003). A hedonic regression of home prices in King County, Washington, using activity-specific accessibility measures. *Transportation Research Board 2003 Annual Meeting*.

[6] Munaiah, V., Maheswari, P., Gangaram, T., Sreenivasulu, K., & Murali, K. (2025). A comprehensive statistical analysis of residential property valuation using multiple linear regression techniques. *Journal of Computational Analysis and Applications (JoCAAA), 34*(7), 462–475.

[7] Chen, Z. (2024). The study on the influencing factors of housing price. In *Proceedings of the 1st International Conference on Innovations in Applied Mathematics, Physics and Astronomy (IAMPA)* (pp. 258–263). SciTePress.

[8] Trojanek, R., Gluszak, M., & Tanas, J. (2018). The effect of urban green spaces on house prices in Warsaw. *International Journal of Strategic Property Management, 22*(5), 358–371.

[9] Sachdeva, M., Fotheringham, S., & Li, Z. (2022). Do places have value? Quantifying the intrinsic value of housing neighborhoods using MGWR. *Journal of Housing Research, 31*(1), 24–52.

[10] Gorjian, M. (2025). Integrating machine learning and hedonic regression for housing price prediction: A systematic international review of model performance and interpretability [MPRA Working paper].