Predicting Housing Prices: The Case of Ankara Housing Market

Duygu Çayan, Özgül Burcu Özdemir Sarı
Affiliation: Gazi University, METU
Email: duygucayan@gazi.edu.tr, bozdemir@metu.edu.tr

Abstract
The housing market is crucial for economies, especially in Türkiye, where economic instability has caused rising housing prices. This study aims to predict trends in the city's rental market using a dataset of 20,000 listings. By using advanced data analysis and machine learning, the research seeks to identify patterns and forecast rental price movements, benefiting households and policymakers. Valuing housing is complex due to its varied characteristics, prompting the use of hedonic pricing theory. This involves a regression model that links housing prices to their features. While traditional linear regression has been used, non-linear machine learning models often provide more accurate predictions. The research offers a framework for predicting housing market trends and fostering equitable urban development, addressing housing needs in Ankara and beyond.

Keywords: Housing Price Prediction, Housing Affordability, Machine Learning, Türkiye, Ankara

1. Introduction

Accurate estimation of house prices is crucial for a broad spectrum of stakeholders across multiple sectors. In this context, not only households but also a diverse range of actors—including landlords, investors, real estate agents, financial institutions and government bodies—require reliable housing price information for their respective purposes. As a critical sector and a major driver of economic activity, the housing market holds significant importance in nearly all economies, and Türkiye is no exception. The country's ongoing economic instability has further heightened the housing sector's relevance.

Given its pivotal role in the Turkish economy, the performance of the housing market is essential not only for ensuring macroeconomic stability but also for addressing issues of housing affordability. However, house prices in Türkiye tend to be volatile, fluctuating in cycles of rise and decline. Despite this, in major cities such as İstanbul, Ankara and İzmir, prices have shown a consistent upward trend. According to a report from the Central Bank of the Republic of Türkiye in April 2025, the Housing Price Index increased by 32.9% in nominal terms. Analysing the housing price indexes for Istanbul, Ankara, and Izmir reveals that in March 2025, prices increased by 2.7%, 2.8%, and 0.2%, respectively, compared to the previous month. Year-over-year comparisons show that index values increased by 34.4%, 41.9%, and 30.9% in Istanbul, Ankara, and Izmir, respectively. In April 2025, the Consumer Price Index recorded an annual change of 37.86%, according to the Turkish Statistical Institute (TURKSTAT, 2025). When comparing nominal housing price growth to consumer inflation, Ankara stands out because its housing prices have increased slightly more than the inflation rate. The widespread price increase across Türkiye, especially in Ankara, has intensified the focus on monitoring changes in housing prices and addressing affordability issues.

This study focuses on Ankara, the nation's capital, as a case study to forecast housing market trends across nine districts. It utilizes a comprehensive dataset comprising 20,000 rental listings collected from a prominent real estate platform. The dataset includes key property features such as price, type, number of rooms, floor area, age, floor level, and geographic location. By applying advanced data analysis and machine learning techniques, this research aims to uncover patterns in rental price behaviour and generate predictive insights. These findings are expected to contribute meaningfully to urban planning strategies and the formulation of effective housing policies.

The remainder of this paper is organized as follows: Section 2 provides a brief review of previous research on housing price estimation, outlining the theoretical and empirical framework for the analysis developed in Rosen (1974). In Section 3, the data and methodology used in this study are described. Section 4 presents and discusses the empirical results. Building on these findings, Section 5 offers a discussion of the results, while Section 6 outlines the policy implications and concludes the paper.

2. Estimation of Housing Prices

Understanding house price trends and producing accurate estimates are also critically important from the perspective of urban planning and development (Kang et al., 2021). However, predicting housing prices is a complex and multifaceted task, as it involves a combination of structural, economic, social, locational, and market-related attributes that must all be considered during valuation processes (Mankad, 2022). This complexity has led many studies on housing prices to rely heavily on hedonic pricing theory (Crespo & Gret-Regamey, 2013).

The initial formal contribution to hedonic price theory was made by Court in 1941, who developed a hedonic price index for automobiles (Bartik, 1987). Two primary approaches have significantly influenced the theoretical development of hedonic prices. The first approach is based on Lancaster's consumer theory (1966), while the second stems from Rosen's model (1974) (Chin & Chau, 2003). Malpezzi (2003) points out that Lancaster advanced microeconomic theory by focusing on the demand side of the market, emphasizing that utility is derived not from the goods themselves but from their characteristics. The Hedonic Price Model, first formalized by Rosen in 1974, conceptualizes a product as a bundle of characteristics, each contributing to its overall price. As Rosen (1974) explains, "Hedonic prices are defined as the implicit prices of attributes, derived from the observed prices of differentiated products and the specific quantities of characteristics associated with them." In simpler terms, a house's price reflects the cumulative implicit value of its individual features, which cannot be priced independently. A hedonic equation is essentially a regression model that relates a house's price to its various characteristics. A differentiated product is defined by a vector of attributes Z = (z1, z2, ... zn), where each zi represents a specific attribute. The market price of a product can thus be expressed as a function of its attributes, p(z) = p (z1, z2, ... zn). This means that the price of the house, p(z), is a function of its characteristics. To estimate the implicit prices of individual attributes, Rosen proposes a regression-based approach where product prices are regressed on their characteristics. This estimation process represents the first step in Rosen's two-stage hedonic pricing model.

The application of the hedonic price model to the housing market is based on several fundamental assumptions. One is the heterogeneity of housing products. Malpezzi (2003) discusses that hedonic models arise due to heterogeneity in the market, including both heterogeneous housing stock and heterogeneous consumers. Housing products are considered heterogeneous due to their differences in locational, structural, or neighbourhood attributes. Each house possesses a unique set of characteristics that influence its value. Also, a house is a long-lasting good that has a significant lifespan, allowing houses of varying ages to coexist within the same market. Furthermore, certain features of a house may be valued differently depending on the geographical location. Homebuyers have distinct utility functions that lead them to value different characteristics across houses differently, in addition to the problem of the presence of different characteristics across houses (Sirmans, Macpherson, and Zietz, 2005).

Another theoretical construction underlying this approach is the "implicit market." The concept of implicit markets refers to the processes of producing, exchanging, and consuming commodities that are primarily (and perhaps exclusively) traded in "bundles" (Sheppard, 1999). This means that implicit markets refer to situations where individual components of a product are not sold separately, but rather as part of a bundle. Some goods, such as automobiles, workers, and houses, can be grouped into a single "market," despite their heterogeneous nature. These markets can't be analysed with traditional economic models because they don't have a single price; instead, they feature a range of prices based on quality and characteristics. The hedonic approach addresses this by suggesting that these goods comprise aggregates of homogeneous parts. While the overall bundle may lack a uniform price, the individual attributes share a more consistent price structure (Sheppard, 1999).

In earlier research, the hedonic pricing model—typically based on linear regression and estimated using Ordinary Least Squares (OLS)—has been widely used to predict housing prices. However, the relationship between property values and their determining factors is often nonlinear and more intricate than linear models can adequately capture (Sheppard, 1999). Key issues such as model specification methods, multicollinearity among independent variables, interactions between predictors, heteroscedasticity, non-linearity of relationships, and the presence of outliers can substantially undermine the model's reliability and effectiveness in accurately assessing housing values (Limsombunchai, 2004). To address rising prediction errors and heteroscedasticity at higher price levels, the hedonic pricing model is often utilized in a semi-logarithmic form (Goodman & Thibodeau, 1995; Sirmans, Macpherson, and Zietz, 2005; Fletcher, Gallimore, and Mangan, 2000). While the semi-logarithmic form helps mitigate heteroscedasticity and capture some non-linearity, several studies have pointed out its limitations. They have limited flexibility in modelling complex, non-linear interactions between housing attributes and prices. In response, various alternative modelling approaches have been developed, many of which offer significant advantages over traditional methods. Recent studies have shown that non-linear models utilizing machine learning techniques provide more accurate and reliable predictions. These models improve upon standard approaches by learning directly from data and continuously refining their predictive capabilities. Among the most commonly used and effective machine learning algorithms for housing price prediction are artificial neural networks, gradient boosting, and random forests.

This study focuses on using random forest machine learning techniques for predicting house prices. Random Forest is a powerful classification algorithm that enhances predictive accuracy by averaging outcomes from multiple decision trees, each trained on different subsets of the dataset (Adetunjia et al., 2022). It uses a process called bagging, or bootstrap aggregating, to train individual trees on random samples. This approach reduces overfitting and improves the robustness of the model. As an ensemble model, Random Forest combines predictions from various trees using majority voting or averaging, resulting in higher accuracy than a single decision tree. This method is particularly effective for large datasets with complex relationships, revealing intricate patterns among the features (Truong, Nguyen, Dang, and Mei, 2020).

3. Data and Methodology

The dataset for this study is sourced from hepsiemlak.com, a prominent real estate platform in Türkiye. It consists of 19,333 house-for-sale advertisements collected in June 2024. This dataset includes key information such as property prices, the number of rooms, the gross area of the house, the floor, the age, housing types (including flat, residence, detached house, etc.), and the neighbourhood in which it is located (Table 1).

Table 1. List of Attributes

Attribute Name	Data Type	Description
price	int64	House price
totalRooms	int64	Number of total rooms (3, 4, etc.)
area	int64	Area of the house (gross)
floor	int64	The floor it is located on
age	int64	Age of the house
housingTypes	int64	House's type (flat, residence, detached house, etc.)
neighborhood	object	The neighborhood in which it is located
district	object	The district in which it is located

The dataset was thoroughly examined for outliers using the Interquartile Range (IQR) method. An outlier, denoted as x, is identified if it falls outside the established boundaries. Specifically, this occurs when the value of x is lower than the lower quartile minus 1.5 times the IQR or higher than the upper quartile plus 1.5 times the IQR. This systematic approach enhances the understanding of the data's distribution and helps maintain the integrity of the analysis. After applying this method to the price column of the dataset, a total of 790 outliers were identified and subsequently removed. This process significantly enhanced the quality of the dataset, which now comprises 18,543 entries.

The data on Ankara's housing stock, collected during the first stage of the study, encompasses a total of 22 districts, as shown in Table 2. However, it is evident that sufficient data could not be obtained, particularly from districts outside the central areas. Consequently, the focus of the study was narrowed to the central districts, which have a more substantial representation of the housing stock. As a result of this process, the dataset was reduced to 16,903.

Table 2. Number of data by districts

District	**Mean**	**Median**	**Count**
AKYURT	2975304.71	2300000.00	105
ALTINDAG	2748210.51	2500000.00	2072
AYAS	2935608.70	2200000.00	23
BALA	3445678.57	2870000.00	28
BEYPAZARI	2369473.68	2250000.00	57
CAMLIDERE	4478571.43	3450000.00	7
CANKAYA	4430305.93	3850000.00	2190
CUBUK	2353887.32	1765000.00	213
ETIMESGUT	3793351.95	3150000.00	1305
GOLBASI	4057836.86	3500000.00	803
GUDUL	5377500.00	4250000.00	8
HAYMANA	2587500.00	2050000.00	4
KAHRAMANKAZAN	3790614.04	3250000.00	57
KECIOREN	2916656.11	2575000.00	2396
KIZILCAHAMAM	4172359.75	4422292.00	24
MAMAK	2509282.10	2250000.00	2320
NALLIHAN	2052571.43	1893000.00	7
POLATLI	2281024.48	1995000.00	1103
PURSAKLAR	3225362.48	2690000.00	2356
SEREFLIKOCHISAR	1611250.00	1325000.00	4
SINCAN	2448722.26	2195000.00	2637
YENIMAHALLE	4033203.88	3650000.00	824

This study employs a comprehensive methodological framework that incorporates both linear regression and random forest models to predict local housing prices. The linear regression model assumes a straightforward linear relationship between the dependent variable and the explanatory variables. In contrast, the random forest model leverages an ensemble learning approach, using multiple decision trees to capture complex interactions and non-linear relationships within the data. The effectiveness of both models is assessed by comparing their prediction accuracy through metrics such as mean absolute error and R-squared values. This comparative analysis aims to enhance our understanding of which model provides superior forecasting capability for local housing markets.

4. Empirical Results

Exploratory data analysis is a crucial preliminary step prior to developing a regression model. Figure 1 illustrates the average housing prices across various districts in Ankara using a grayscale colour scheme to represent price levels. Darker shades correspond to higher average prices, while lighter shades indicate lower ones. According to the map, Çankaya, Gölbaşı, and Yenimahalle stand out with significantly higher average housing prices compared to other districts. In contrast, Etimesgut, Pursaklar, and Keçiören fall into the mid-priced category, represented by medium gray areas. The lightest shades mark the most affordable districts, which include Altındağ, Mamak, and Sincan.

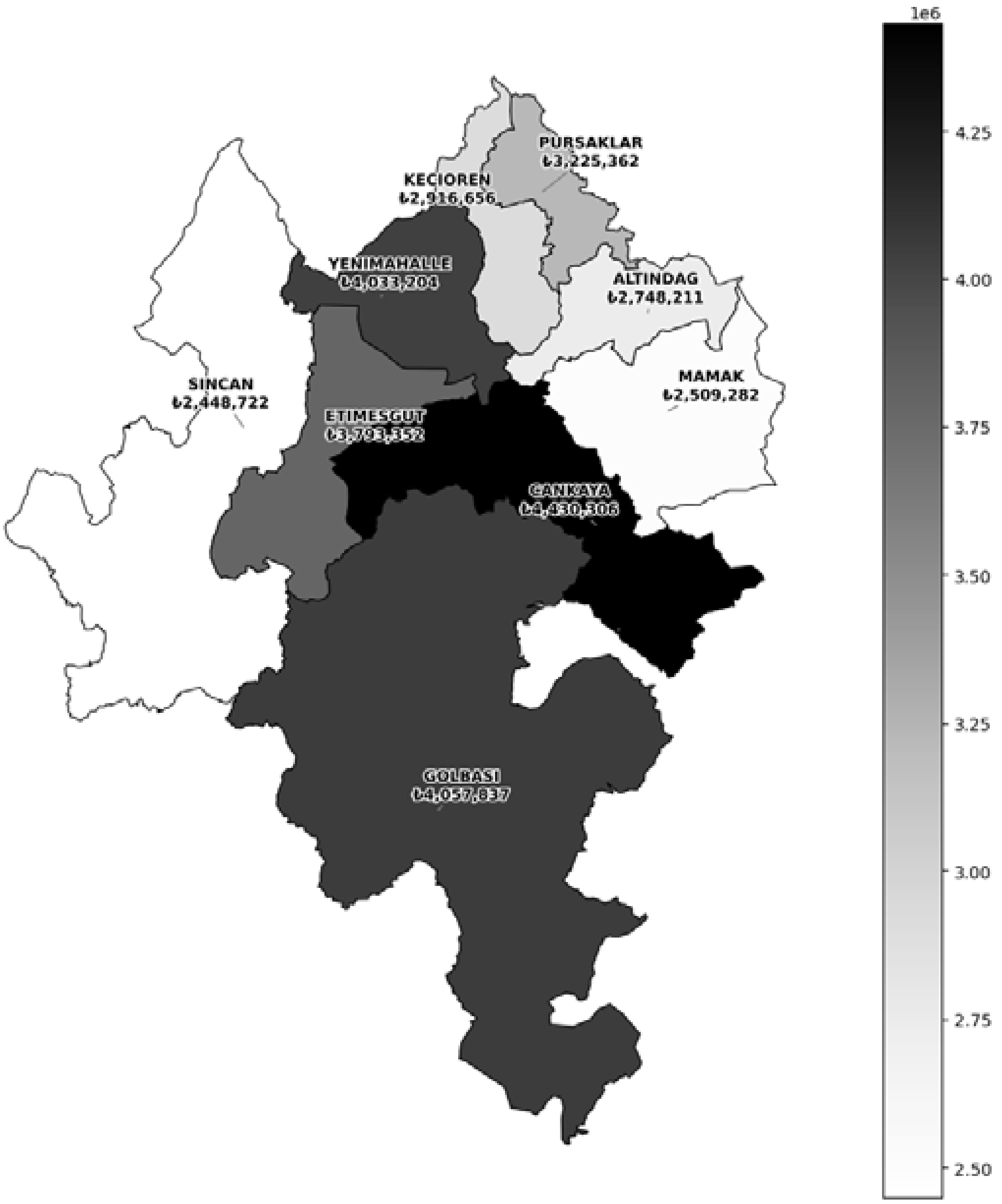

Fig. 1. Average Price by Districts

Figures 2, 3, 4, and 5 compare the distribution of price with its log-transformed counterpart, log Price. The histograms in Figures 3 and 4 reveal a pronounced right skew, while the Q-Q plot shows significant deviations from a normal distribution, especially in the tails. This indicates that the price distribution is not normal. In contrast, the histograms in Figures 5 and 6 appear more symmetrical and bell-shaped. The Q-Q plot for the log-transformed prices aligns much more closely with the theoretical normal line, exhibiting only minor deviations in the tails. This suggests that log-transformed housing prices (log Price) is approximately normally distributed.

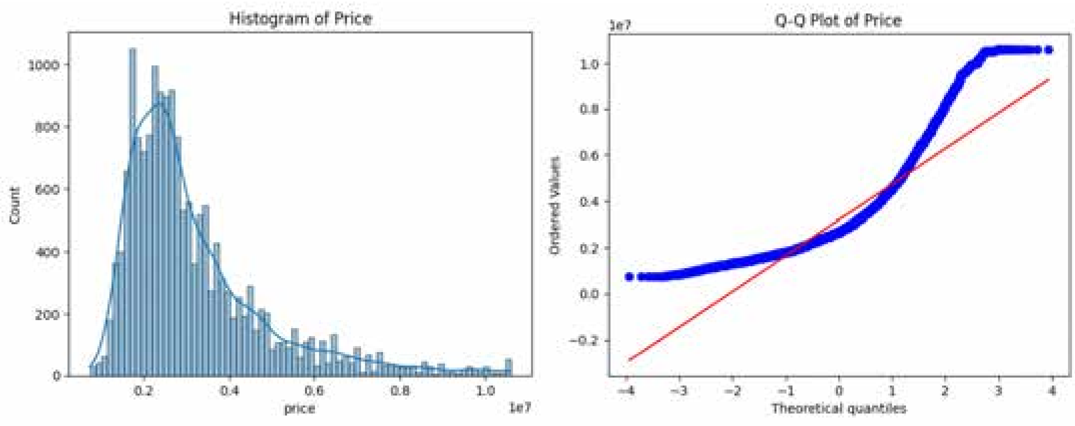

Fig. 2 and 3. Histogram and Q-Q Plot of Price

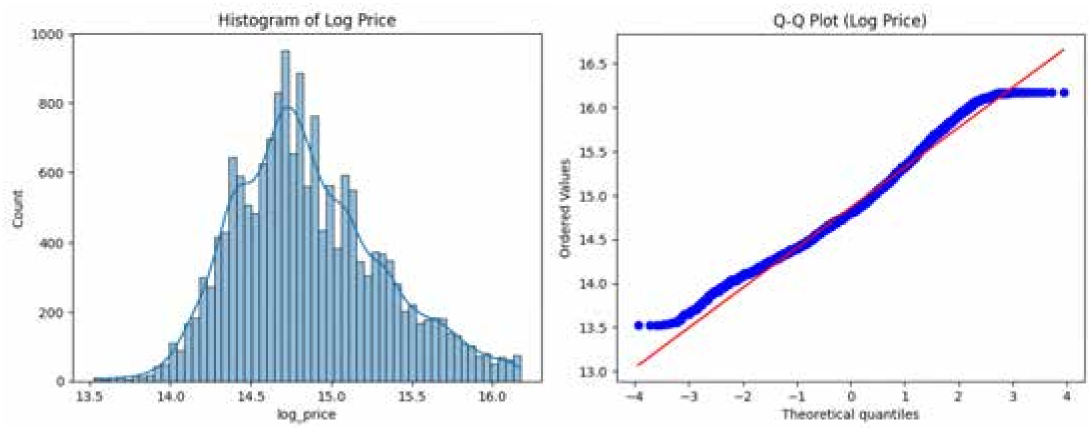

Fig. 4 and Fig. 5. Histogram and Q-Q Plot of Log Price

Figure 6 presents a box plot illustrating the distribution of log-transformed housing prices (log Price) across various districts in Ankara. The x-axis represents the log Price, while the y-axis lists the districts. Each box depicts the interquartile range (IQR), which captures the middle 50% of the data. The horizontal line within each box indicates the median log Price for the corresponding district. The whiskers extend to show the range of values excluding outliers, which are represented by individual circles.

According to the plot, there are significant variations in housing prices across 9 districts. Çankaya emerges as the most expensive district on average, characterized by a wide price range and a substantial number of high-end outliers. Sincan, Mamak, and Altındağ appear as the most affordable districts, with lower median log Prices. Etimesgut and Yenimahalle fall into the mid-range category, exhibiting moderate price levels along with some variability. Districts such as Pursaklar and Altındağ display tighter distributions, suggesting relatively more homogeneous housing markets in those areas.

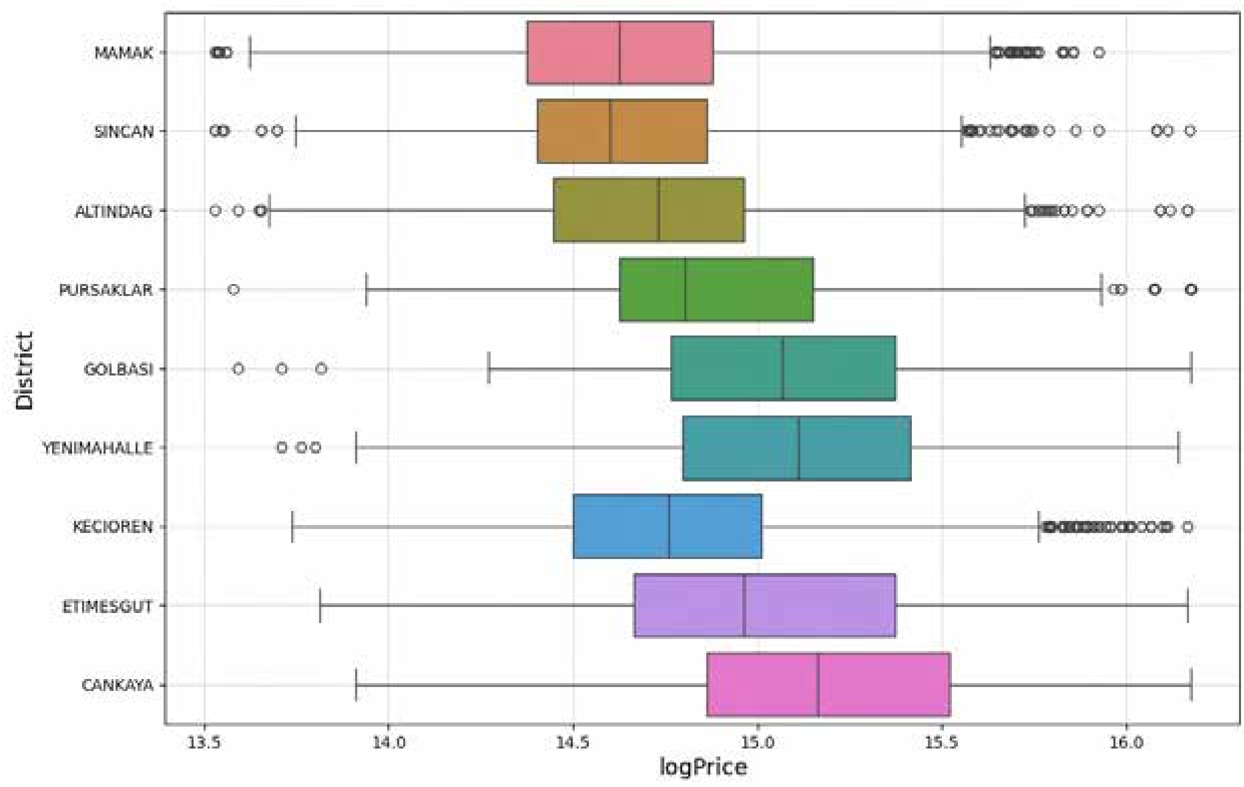

Fig. 6. District Log Transformation of price

Figures 7 and 8 illustrate the overall distribution of unit prices, which follows a right-skewed pattern. The average unit price is approximately ₺23,300, while the median is ₺20,562. This difference suggests that some high-priced properties are elevating the average price. Among the districts, Gölbaşı and Çankaya have the highest average unit prices, indicating they feature more upscale or premium housing markets. In contrast, Sincan, Mamak, and Altındağ are among the most affordable districts. Additionally, Gölbaşı exhibits the greatest price variability, indicating a wide range of property types, while Sincan shows more consistent pricing. These trends underscore significant spatial disparities in housing costs across the region.

	count	mean	std	min	25%	50%	75%	max
unit_price	16903.0	23304.551620400423	10621.32212196031	1664.5833333333333	16581.14035087719	20562.5	26600.0	211111.11111111112

Fig. 7. Descriptive Statistics for Unit Price

	count	mean	std	min	25%	50%	75%	max
ALTINDAG	2072.0	20429.4477690962	7387.4215255031395	3200.0	15500.0	19090.909090909092	23500.0	108333.33333333333
CANKAYA	2190.0	33475.56164597417	14128.748510837346	10000.0	24044.471153846156	30000.0	39800.0	211111.11111111112
ETIMESGUT	1305.0	26583.541578259985	9818.183830033804	7996.666666666667	19150.0	24305.555555555555	32954.545454545456	66250.0
GOLBASI	803.0	37239.66147111162	15597.817484212028	1664.5833333333333	25912.962962962964	35000.0	45313.725490196084	105555.55555555556
KECIOREN	2396.0	20424.769618286788	6615.683964992601	7867.64705882353	16067.994505494506	19449.928977272728	22970.299145299145	66666.66666666667
MAMAK	2320.0	18757.841455021917	6225.121634379179	7796.0	14800.0	17500.0	21111.11111111111	63425.92592592593
PURSAKLAR	2356.0	21417.222962977292	6049.4035584624235	4625.0	17619.04761904762	20454.545454545456	24000.0	74561.40350877192
SINCAN	2637.0	18046.518339128776	5104.756206356405	2343.75	14666.666666666666	17158.333333333332	20400.0	66000.0
YENIMAHALLE	824.0	28127.407760100305	11265.089033059183	9000.0	20996.913580246914	25714.285714285714	32222.222222222223	100000.0

Fig. 8. Descriptive Statistics for Unit Price by Districts

Before constructing the models, it's crucial to preprocess the data to improve the models' ability to learn patterns effectively. Specifically, numerical values were standardized to bring them into a comparable range, which aids in achieving better convergence during training. Meanwhile, categorical values, such as those in the district column, were transformed using one-hot encoding, which generates binary variables for each category.

The correlation matrix (Fig. 9) offers valuable insights into the relationships between various housing characteristics and their impact on pricing. It shows that price correlates moderately to strongly with area (m²) at 0.58 and with total rooms at 0.45. This correlation indicates that larger homes tend to command higher prices in the housing market. Additionally, there is a significant negative correlation of -0.24 between price and building age, suggesting that newer properties generally have higher prices. This trend can be attributed to the modern amenities and superior construction standards often found in newer homes, making them more attractive to buyers. Moreover, the high correlation of 0.82 between area (m²) and total rooms raises concerns about multicollinearity, which should be carefully considered in predictive modelling. Multicollinearity can obscure the true relationships between independent variables and the dependent variable, leading to less reliable predictions. Upon further examination of the data, it becomes clear that specific property types display distinct pricing trends. For example, villas show a positive association with price, likely due to their luxury appeal and larger lot sizes. In contrast, flats exhibit a negative correlation with pricing, which may reflect their smaller size or less desirable features compared to other property types. Locational differences also play a significant role in pricing dynamics. Certain districts, such as Çankaya, are positively correlated with higher prices, presumably because of their central location, better amenities, and perceived prestige. Conversely, areas like Sincan and Altındağ show negative relationships with price, which may be linked to socioeconomic factors, market demand, or varying levels of property development in those districts. This pricing disparity highlights the importance of location in housing prices.

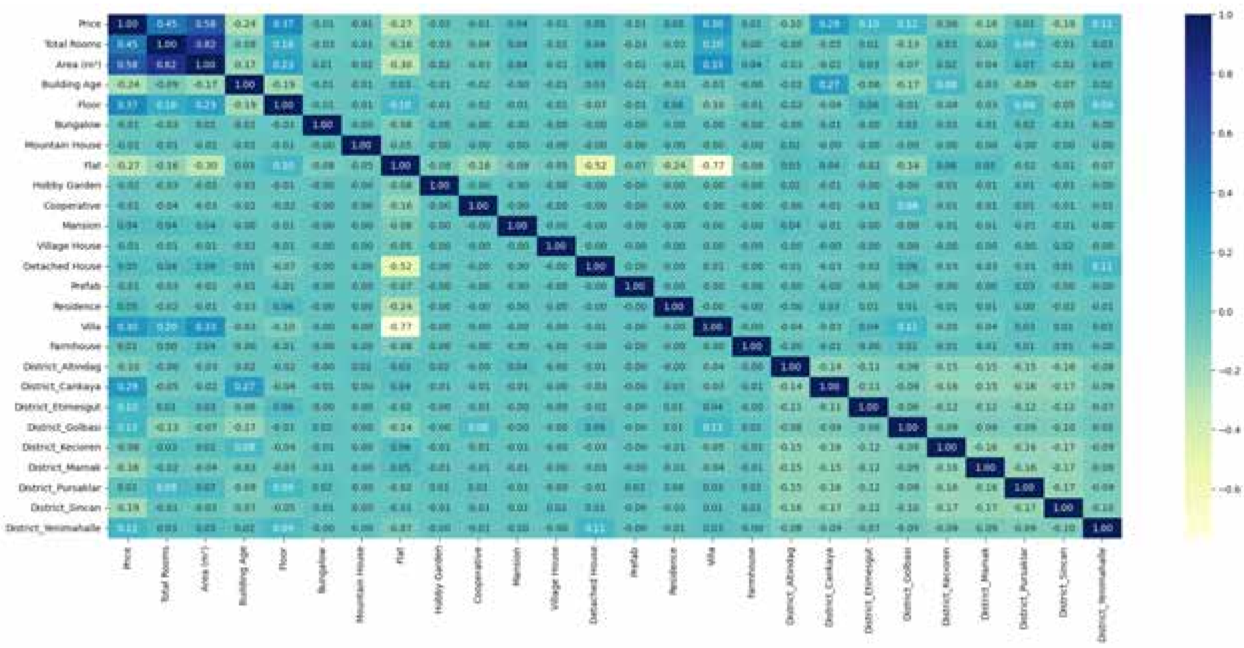

Fig. 9. Correlation Matrix

4.1. Baseline Model Results

In the study, a baseline model was developed to compare linear and random forest models. This baseline model predicts housing prices based solely on the average price per neighbourhood. The results of the model are presented in Table 3. It has a Mean Squared Error (MSE) of approximately 1.7 trillion, indicating a substantial average error in squared units. The Root Mean Squared Error (RMSE) is around 1,310,163, which represents the average prediction error in actual price units—this is quite high and suggests that the model lacks precision. The R-squared (R^2) value is 0.367, meaning the model explains only about 39.9% of the variance in housing prices.

This implies that while neighbourhood averages provide some predictive power, the baseline model is relatively weak, and more sophisticated modelling, such as using multiple features, is necessary for improved accuracy.

Table 3. Baseline Model Results

Mean Squared Error	1,716,527,408,237.00
Root Mean Squared Error	1,310,163.00
R-squared	0.3998

Figure 10 shows predicted housing prices versus actual prices. The red dashed line indicates where predictions would perfectly align with actual values. Many points are close to this line, suggesting reasonably accurate predictions, but there's more variability at higher price ranges. This indicates the model is less precise with expensive properties, reflecting some significant under- and overestimations. While it captures the general trend, it shows room for improvement through additional features or more complex modelling techniques.

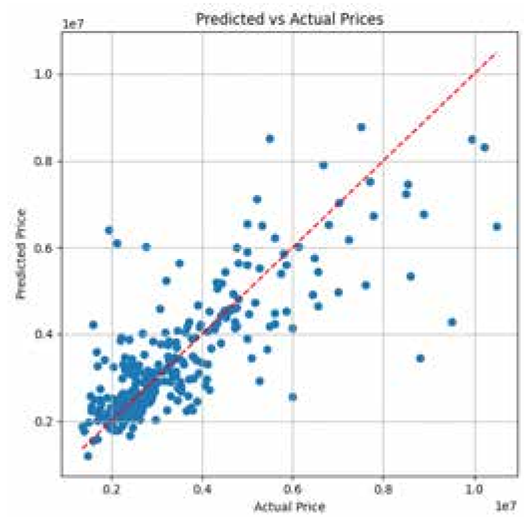

Fig. 10. Baseline Model Predicted and Actual Price

4.2. Linear Regression Results

In the second phase of the study, housing prices were assessed using a linear regression model. The outcomes of the model are outlined in Table 4. The Linear Regression model exhibits a moderate level of predictive performance, with an R-squared value of 0.6108. This indicates that approximately 61% of the variability in housing prices can be explained by the model's features. However, the high Mean Squared Error (over 1.1 trillion) and a Root Mean Squared Error of about 1,045,316 suggest that the model's predictions significantly deviate from the actual prices in absolute terms. This level of error may be considerable, depending on the typical price range in the dataset. Overall, while the model identifies some meaningful trends, its precision is limited, indicating significant room for improvement. This could potentially be achieved through the use of more advanced modeling techniques or enhanced feature engineering.

Table 4. Linear Regression Results

Mean Squared Error	1,092,686,523,151.36
Root Mean Squared Error	1,045,316.47
R-squared	0.6108

Figure 11 illustrates the predicted housing prices compared to the actual prices generated by the Linear Regression model. There is a general upward trend, indicating that the model captures some relationship between the features and the target variable. However, the predictions tend to cluster within a narrower price range. Notably, the model underestimates the prices of high-end properties while overestimating some lower-priced ones. This is evident as many data points significantly deviate from the red dashed line, which signifies perfect prediction. The scatter of points, particularly at both the high and low ends of the price spectrum, visually reinforces the earlier numerical evaluation, highlighting the model's difficulty in accurately capturing the full variability of housing prices.

Fig. 11. Linear Regression Predicted and Actual Price

Applying a logarithmic transformation to the target variable (price) in the linear regression model resulted in a slight improvement in explanatory power, with the R^2 value increasing from 0.6108 to 0.6168. More significantly, the Root Mean Squared Error (RMSE) experienced a dramatic decrease from over 1 million in raw prices to just 0.28 on a log scale. This indicates a substantial reduction in relative error. The findings suggest that the model is more effective at predicting proportional (percentage) differences in price rather than absolute price values. Additionally, log transformation stabilizes variance and diminishes the impact of extreme outliers, leading to more consistent predictions across various price ranges. Although the log price transformation slightly improved the model's predictive accuracy (as reflected by a higher R^2 value), converting the predictions back to actual price values resulted in significantly inflated estimates. This suggests that while the log transformation helps with model fitting, it may distort the predicted values when inverse-transformed, especially in the presence of skewed data or outliers. For this reason, the actual price value was used as the dependent variable in the study instead of the log transformation of the price.

4.3. Random Forest Results

In the third phase of the study, housing prices were evaluated using a random forest model (see Table 5). The Random Forest Regressor model, after undergoing hyperparameter tuning, achieves an R^2 of 0.7233. This indicates that the model explains approximately 72% of the variance in housing prices, which is a significant improvement over linear regression. The model's Root Mean Squared Error (RMSE) is around 881,343 reflecting a lower average prediction error and demonstrating its ability to effectively manage the complexity and nonlinearity of the housing data. When compared to the untuned version, which had an R^2 of 0.7227 and an RMSE of approximately 882,200 the tuning process has resulted in a modest improvement. Overall, this tuned Random Forest model provides robust and reliable results for predicting property prices.

Table 5. Random Forest Results

Mean Squared Error	776,765,523,406.89
Root Mean Squared Error	881,343.02
R-squared	0.7233

Compared to the linear regression model, the scatter plot for the Random Forest Regression (Figure 12) shows a more accurate alignment of predicted prices with actual prices, particularly for mid- to high-range properties. While both models display some spread, the Random Forest's predictions are generally closer to the diagonal reference line, indicating improved accuracy. It captures nonlinear relationships better, reducing systematic underestimation of expensive properties seen in the linear model. Although slight over- and under-predictions remain, especially at the extremes, this plot visually confirms that Random Forest provides a more reliable fit than linear regression in modelling complex housing price patterns.

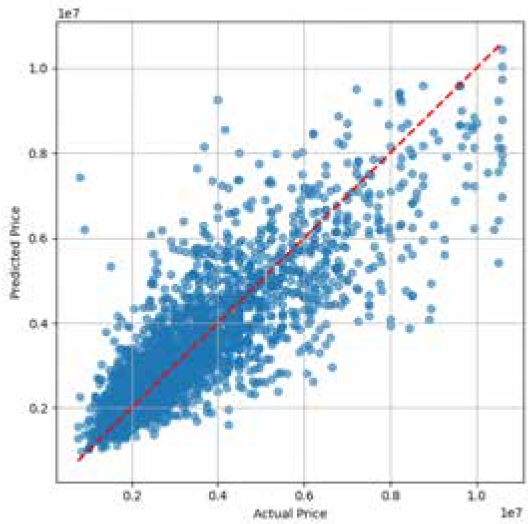

Fig. 12. Random Forest Predicted and Actual Price

5. Discussion

This study aims to achieve a more precise estimation of housing prices by utilizing advanced analytical methods. To accomplish this, it compares both linear and non-linear specifications of the hedonic price model, a widely recognized framework in housing market research. Table 6 compares actual housing prices with predictions made by Linear Regression and Random Forest models across the 15 most expensive neighbourhoods. The Linear Regression model tends to underestimate housing prices, particularly in high-value areas such as Fatih Sultan, Mutlukent, and Çukurambar. This indicates the model's limitations in capturing non-linear patterns and extreme values. In contrast, the Random Forest model provides more accurate and stable predictions, often aligning closely with actual prices and even slightly overestimating in some cases, such as in Koru. This suggests that the Random Forest model is better suited for capturing the complex relationships present in housing data and performs more robustly in high-price segments. However, both models exhibit certain deviations from actual prices, which highlights opportunities for further improvement.

Table 6. The comparison Linear Regression and Random Forest Prediction with Actual Price

Neighborhood	Actual Price	Linear Regression Prediction	Random Forest Prediction
Fatih Sultan	9,850,000	6,378,664	7,845,811
Mutlukent	8,600,000	5,563,571	6,469,399
Çukurambar	8,525,000	5,180,941	5,590,277
Üniversiteler	8,500,000	4,030,711	4,646,521
Emniyet	8,475,000	6,275,043	6,284,633
Koru	8,450,000	4,414,746	3,867,666
Kızılırmak	7,833,333	5,140,973	6,797,407
Alacaatlı	7,815,000	5,886,321	6,887,986
Koparan	7,700,000	10,320,204	9,194,666
Bademlidere	7,606,250	7,221,187	7,895,229
Yaşamkent	7,180,000	7,017,621	7,152,333
Bağlıca	6,740,580	4,632,440	5,920,819
Malıköy	6,649,166	6,962,850	6,251,322
Bahçelievler	6,587,500	4,444,840	5,760,515
Beştepe	6,500,000	4,066,717	5,365,666

Based on the results of the predictive modelling, housing affordability in Ankara varies significantly across neighbourhoods. Certain neighbourhoods, such as Fatih Sultan, Mutlukent, Üniversiteler, and Çukurambar, consistently show high actual prices, indicating low affordability, particularly since the model tends to underestimate these values. This underestimation may be due to unobserved amenities or factors that contribute to desirability, which are not fully captured by the model's features. Additionally, the substantial prediction errors in high-priced neighbourhoods suggest that affordability challenges in these areas may be more severe than initially indicated by the data. In contrast, the model performs better in moderately priced neighbourhoods, which may imply more predictable and accessible housing conditions in those locations. To enhance the analysis and provide a more comprehensive understanding of affordability across the city, incorporating socioeconomic indicators such as household income, employment rates, and demographic characteristics would be beneficial.

The analysis indicates that the neighbourhoods with the highest housing prices—Kızılırmak, Çukurambar, Üniversiteler, Mutlukent, Yaşamkent, Koru, Alacaatlı, Fatih Sultan, and Bağlıca—are clustered along the Eskişehir Road, to the west of the city centre. It can also be inferred that the properties in the adjacent neighbourhoods of Çayyolu and Beytepe are unaffordable. Furthermore, high housing prices are observed in the neighbourhoods located west of the Konya Road, which serves as one of Ankara's main transportation axes. These neighbourhoods include Beştepe, Emniyet, İşçi Blokları, and Çiğdem. In Gölbaşı, the neighbourhoods of Ballıkpınar, Hacılar, Tulumtaş, and Koparan also experience elevated prices. This trend can be attributed to the fact that the housing typologies in these areas are primarily detached homes.

The analysis highlights a concerning trend of increasing housing unaffordability in Ankara, especially in neighbourhoods along the Eskişehir and Konya Road corridors. These areas, known for their high-income residents and mainly planned developments with detached single-family homes, have seen substantial increases in property prices in recent years. As a result, housing in these districts has become prohibitively expensive for middle- and lower-income residents. This ongoing rise in housing costs worsens spatial inequality throughout the city. Not only is housing affordability compromised, but the diverse socio-economic backgrounds that once characterized these neighbourhoods are also increasingly at risk. The current scenario in Ankara highlights a significant trend of urban segregation, where many individuals find the aspiration of owning a home increasingly out of reach. This reality not only deepens existing social divides but also inhibits community cohesion, ultimately impacting the overall cohesion within the city.

6. Conclusion

Given the increasing economic pressures and rising housing prices in Turkey, especially in Ankara, it has become crucial to develop accurate and effective methods for predicting housing market trends. To monitor housing market dynamics, the hedonic price model has long been used to monitor housing market dynamics. However, this model has some structural limitations. Predicting housing prices is challenging. The Central Bank of Türkiye has published a Residential Property Price Index monthly since 2010 to monitor price movements, but it only includes dwelling units valued for housing loans, representing just 35–40% of the market (Özdemir Sarı, 2019). This limits its effectiveness in capturing the overall housing market dynamics. Traditional hedonic price models often use linear regression methods, which fail to account for the complex, non-linear relationships between housing prices and their attributes. This study proposes an innovative approach aimed at refining housing price estimates by effectively addressing the inherent limitations of traditional hedonic models and significantly enhancing the representativeness of the data utilized. Through a comprehensive analysis that compares linear regression and random forest methodologies, the findings indicate that the random forest model consistently yields superior predictive accuracy and reliability.

The findings highlight growing concerns about housing affordability in Ankara, revealing significant disparities between neighbourhoods. Wealthier districts along the Eskişehir and Konya Road corridors—such as Fatih Sultan, Çukurambar, Mutlukent, and Alacaatlı—consistently show high property prices, which predictive models often underestimate. This discrepancy indicates that these areas likely possess unmeasured qualities, such as better amenities or prestige, that structural features alone do not capture. Consequently, affordability in these neighbourhoods is probably even lower than suggested, creating barriers for middle- and low-income households. Additionally, the concentration of high-priced, low-affordability neighbourhoods in the western and southwestern parts of the city reveals a trend of spatial and socioeconomic segregation. The increasing exclusivity of these areas, combined with rising prediction errors in high value zones, suggests a widening gap in access to housing. To fully understand and address this issue, future modelling efforts should incorporate socioeconomic indicators and neighbourhood-level attributes. Doing so would help capture the complex nature of urban housing markets and support more informed urban planning and policy decisions aimed at improving housing accessibility and equity.

The current analysis relies solely on structural characteristics of housing and neighbourhood as a binary variable. Incorporating additional location-specific and neighbourhood attributes, such as proximity to amenities, schools, public transportation, crime rates, and green spaces, has the potential to significantly enhance the accuracy of the predictive model. Urban housing markets exhibit considerable complexity, rendering them ineffective as a singular, homogeneous market. This complexity arises from a variety of factors, including socioeconomic diversity, varying demand and supply dynamics, and distinct local regulations. As a result, making predictions tailored to specific sub-markets within the urban landscape can lead to improved prediction rates and more reliable insights into housing trends and pricing fluctuations.

References

Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2022). House price prediction using random forest machine learning technique. Procedia Computer Science, 199, 806-813.

Bartik, T. J. (1987). The estimation of demand parameters in hedonic price models. Journal of political Economy, 95(1), 81-88.

Central Bank of the Republic of Türkiye. (2025). Residential Property Price Index: April 2025. Data Governance and

Statistics Department, Surveys and Indices Division. [Online] available at: https://www.tcmb.gov.tr/wps/wcm/connect/EN/TCMB+EN/Main+Menu/Statistics/Real+Sector+Statistics/Residential+Property+Price+Index/ [Accessed 7 Jun. 2025].

Chau, K. W., & Chin, T. L. (2003). A critical review of literature on the hedonic price model. International Journal for Housing Science and its applications, 27(2), 145-165. ion of housing submarkets. Environment and Planning A, 33(12), 2235-2253

Crespo, R. and Gret-Regamey, A. (2013) Local hedonic house-price modelling for urban planners: advantages of using local regression techniques. Environment and Planning B: Planning and Design, 40, pp.664–682.

Fletcher, M., Gallimore, P., & Mangan, J. (2000). Heteroscedasticity in hedonic house price models. Journal of Property Research, 17(2), 93-108.

Goodman, A. C., & Thibodeau, T. G. (1995). Age-related heteroskedasticity in hedonic house price equations. Journal of Housing Research, 25-42.

Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F. and Ratti, C. (2021) Understanding house price appreciation using multi-source big geo-data and machine learning. Land Use Policy, 111, p.104919.

Limsombunchao, V. (2004). House price prediction: hedonic price model vs. artificial neural network.

Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. Housing economics and public policy, 1, 67-89.

Mankad, M.D. (2022) Comparing OLS based hedonic model and ANN in house price estimation using relative location. Spatial Information Research, 30 (1), pp.107–116.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. Journal of political economy, 82(1), 34-55.

Sheppard, S. (1999). Hedonic analysis of housing markets. Handbook of regional and urban economics, 3, 1595-1635.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The composition of hedonic pricing models. Journal of real estate literature, 13(1), 1-44.

Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing price prediction via improved machine learning techniques. Procedia Computer Science, 174, 433-442.

TURKSAT (2025). Consumer Price Index, April 2025. [Online] available at: https://data.tuik.gov.tr/Bulten/Index?p=Tuketici-Fiyat-Endeksi-Nisan-2025-54179 [Accessed 7 Jun. 2025].

Özdemir Sarı, Ö. B. (2019). Redefining the housing challenges in Turkey: An urban planning perspective. Urban and regional planning in Turkey, 167 184.