



Research Article

Sale price classification models for real estate appraisal

Merve Kuru Erdem ¹, Onur Yiğit Erdem ², Gulben Calis* ¹

¹ Ege University, Department of Civil Engineering, Izmir (Turkey) mervekuru25@gmail.com; gulben.calis@ege.edu.tr

² Izmir Metropolitan Municipality, Izmir (Turkey), onur_erdem_90@hotmail.com

*Correspondence: gulben.calis@ege.edu.tr

Received: 11.01.2021; **Accepted:** 02.11.2021; **Published:** 31.12.2021

Citation: Kuru, M., Erdem, O. and Calis, G. (2021). Sale price classification models for real estate appraisal. *Revista de la Construcción. Journal of Construction*, 20(3), 440-451. <https://doi.org/10.7764/RDLC.20.3.440>.

Abstract: This study aims to determine the parameters that are effective on sale prices and to obtain the functions that determine the appropriate sale price ranges of real estates. In this context, a total of 138 real estates, which are located in Bayraklı district of Izmir, Turkey and that were for sale between April and June 2019, were investigated. The effects of 17 parameters on the sale price of the real estates were examined through statistical analysis. Thirteen parameters that have been determined to be effective have been used in developing the distinctive functions that decide the sale price ranges of real estates. The results show that parameters such as real estate's area, age, furniture status, central heating system, playground, pool and gym are statistically significant on the sale price of the real estates. In addition, the functions obtained by using these parameters classified 78.3% of real estate sale prices in the correct range.

Keywords: real estate appraisal, sale price function, discriminant analysis, classification.

1. Introduction

Real estates are considered to be profitable resources regarding investment purposes, and, thus, it is of great importance to accurately determine the value of real estate in trading or exchanges (Erdem, 2017). In the literature, various parameters are reported to affect the sale prices of real estates. Afşar (2017) states that area, central heating, elevator, built-in kitchen, number of bathrooms, number of rooms, parking lot, en-suite bathroom and being on the first floor are the most effective parameters on the sale price. Daşkıran (2015) concluded that the floor number, elevators, rooms and bathrooms, the heating system and the distance of the real estate to educational institutions, health institutions and the city center significantly affected the sale prices. Çiçek and Hatırlı (2015) found that the most important factors affecting the real estate prices in Isparta, Turkey are neighborhood, age, area, number of rooms, central heating, parking lot, and the distance to the city center. Ayan and Erkin (2014) found that air quality of the site and school district affected the sale price of flats in Izmit, Turkey. Yayar and Karaca (2014) concluded that the number of bathrooms and elevators, central heating and being on the boulevard affected the sale prices positively. Koramaz and Dokmeci (2012) concluded that spatial parameters such as distance to the central business district, sub-centers, main transport arteries and beach can explain half of the change in sale prices in Istanbul, Turkey. Marmolejo-Duarte (2017) showed that urban centrality impacts sale prices in Barcelona, Spain. Kangallı Uyar and Yayla, (2016) have found that Bosphorus view caused the most increase in the prices of real estate in Istanbul with 34.04%. Ferlan et al. (2017) showed that the distance from the city center is the most important factor on the selling price in Slovenia. Chiarazzo et al. (2014) showed that the most important parameters affecting prices in the Taranto region of Italy are the distance to the beach and the presence of the garden or the terrace. Furthermore, Uğur and Leblebici (2019) showed that energy efficiency

and water saving parameters increased the sale prices. Marmolejo-Duarte and Chen (2019) also showed that energy efficiency has impact on sale prices.

In the literature, there are various models to estimate real estate sale price. Regression analysis attracted attention among researcher for determining real estate sale prices. Bin et al. (2017) proposed a regression model based on artificial neural networks that estimates the average selling price of a real estate, taking into account the characteristics, position and market price trend of real estate. Bhagat et al. (2016) predicted real estate sale prices with a model based on linear regression by analyzing previous market trends, price ranges and developments. Manjula et al. (2017) estimated the sale prices of real estates by using multivariate regression analysis, taking into account area, age, number of bedrooms, floor number, date of sale, sea view, postal code, latitude, and longitude. Sarip et al. (2016) estimated real estate sale prices with fuzzy regression model based on land area, building area, number of bedrooms and bathrooms, age, neighborhood, and in-built-furniture quality. In addition, different hedonic pricing models were proposed for determining the sale prices. Can (1992) proposed four alternative modeling approaches for the functional specification of the hedonic price function which corresponds to how locational effects are treated in the housing price determination process.

The results indicated that models which incorporate both types of externalities are superior to the specifications in the mainstream literature in which only neighborhood effects are considered. Cao et al. (2019) developed three hedonic pricing models, including an ordinary least squares model, a Euclidean distance-based geographically weighted regression model, and a travel time-based model supported by a big data set of millions of smartcard transactions, to investigate the spatial variation of Housing Development Board public housing resale prices in Singapore. The results show that the age and the floor area of the housing units, the distance to the nearest park, the distance to the central business district, and the distance to the nearest Mass Rapid Transit station parameters significantly affect public housing resale prices. Recently, Wang et al. (2019) used time series analysis for determining sale prices and developed an estimation model based on deep learning and ARIMA model, taking into account area, age, neighborhood, orientation, number of rooms, bedrooms and bathrooms, presence of an elevator, accessibility to schools and the subway.

The literature review show that determining the exact sale price of a real estate is challenging. In order to determine the sale price of a real estate, first, the most effective parameters and their impacts on the sale prices need to be identified. However, the relationship between parameters is complex and dynamic. Furthermore, the sale price of a real estate is highly impacted by intangible assets. The Dictionary of Real Estate Appraisal defines intangible property as nonphysical assets, including but not limited to franchises, trademarks, patents, copyrights, goodwill, equities, securities, and contracts as distinguished from physical assets such as facilities and equipment (Appraisal Institute, 2015). If the real estate cannot be sold without the intangible, then the intangible is probably not an asset on its own but, instead, part of the real property. For example, the view of a real estate is intangible in nature and it cannot be sold without the real estate, nor can the real estate be sold without it. Accordingly, the contribution of the view to the overall value of real estate cannot be determined objectively. The fact that intangible assets are appraised subjectively, utilization of sale price estimation models for real estates might be considered as an unrealistic approach. Therefore, this study aims to:

- Determine which parameters related to the properties of real estates are effective in the sale prices
- obtain functions that determine suitable sale price ranges for real estates using effective parameters, and
- identify the degree of impact of parameters related to real estate properties

In this regard, a total of 138 real estates in Bayraklı district of Izmir and 17 associated parameters were analyzed. The effects of these parameters on real estate sale prices and classification functions based on the parameters were developed via statistical methods. The following sections introduce the field study, explain the methodology, and present the results, discussion and conclusion.

2. Field study

Within the scope of the study, a total of 138 properties in high-rise buildings that are listed for sale between April and July 2019 in Bayraklı district of Izmir province were investigated (Sahibinden.com, 2019). A total of 17 parameters were selected and grouped for analyses. The parameter groups as well as the average sale price of the real estates per each group are presented in Table 1.

Table 1. Averages of real estate sale prices according to the parameter groups.

| | Data Number | Mean Price (USD) |
|--------------------------------|-------------|------------------|
| Property type | | |
| Flat in a low-rise apartment | 89 | 90.042 |
| Flat in a high-rise apartment | 49 | 157.748 |
| Listed by | | |
| Real estate agency | 117 | 118.958 |
| Owner | 21 | 86.918 |
| Floor number | | |
| 1-3 | 30 | 83.174 |
| 4-6 | 27 | 107.797 |
| 7-9 | 25 | 109.722 |
| 10-12 | 25 | 110.068 |
| 13-15 | 15 | 132.818 |
| 16-18 | 16 | 177.924 |
| Move-in status | | |
| Empty | 70 | 119.654 |
| On rent | 30 | 119.442 |
| Occupied by the property owner | 38 | 99.588 |
| Building status | | |
| New construction | 25 | 102.864 |
| Existing | 113 | 116.564 |
| Furnished | | |
| Present | 19 | 138.423 |
| Not Present | 119 | 110.196 |
| Balcony | | |
| Present | 128 | 113.015 |
| Not Present | 10 | 127.751 |
| View | | |
| City | 91 | 92.409 |
| Sea | 47 | 156.046 |
| Building age | | |
| 0-5 | 78 | 134.679 |
| 6-10 | 12 | 95.368 |
| 11-15 | 20 | 84.708 |
| 16-20 | 28 | 79.940 |
| Number of rooms | | |

| | | |
|------------------------------|----|---------|
| 1 | 22 | 103.961 |
| 2 | 23 | 109.408 |
| 3 | 93 | 142.665 |
| Number of bathrooms | | |
| 1 | 65 | 111.447 |
| 2 | 73 | 116.429 |
| Gross area (m ²) | | |
| 50-100 | 16 | 90.257 |
| 101-150 | 80 | 102.752 |
| 151-200 | 42 | 144.741 |
| Net area (m ²) | | |
| 50-100 | 36 | 111.276 |
| 101-150 | 90 | 105.331 |
| 151-200 | 12 | 188.141 |
| Air conditioning | | |
| Central | 77 | 133.659 |
| Individual | 61 | 89.371 |
| Playground | | |
| Present | 64 | 93.357 |
| Close by | 74 | 132.008 |
| Pool | | |
| Present | 68 | 139.346 |
| Not present | 70 | 89.541 |
| Gym | | |
| Present | 79 | 135.865 |
| Not present | 59 | 84.916 |

3. Method

In this study, the data was analyzed in two steps and the SPSS program was used to investigate whether the parameters of 138 real estates have an effect on the sale price of the real estate (SPSS software, 2017). In the first step of the analysis, variance analysis was carried out to examine whether the selected parameters have an effect on real estate sale prices. In the second step, the differential function, also known as discriminant analysis, was conducted in order to create the classification functions of real estate sale prices based on the parameters and determine the priority order of these parameters in the functions.

In the first step, statistical significance between the averages of different real estate sale price groups defined according to the parameters was determined. In other words, tests were conducted to find out whether the variance between the groups is due to the selected parameter or not. If the distribution of the dependent variable according to the selected parameter levels is suitable for the normal distribution, then parametric tests are used, otherwise non-parametric tests are used. It should be noted that Kolmogorov-Smirnov or Shapiro-Wilk normality tests are used to find conformity to normal distribution. In addition, Levene's test is used to control homogeneity of group variances.

If the selected parameter has 2 levels and at the same time:

- If the distribution of the groups of the dependent variable generated according to these levels is suitable for normal distribution and the variances of the groups are homogeneous, then the independent sample t-test accepting the equality of the variances,
- If the distribution of the groups of the dependent variable generated according to these levels is suitable for normal distribution and the variances of the groups are not homogeneous, then the independent sample t-test that does not accept the equality of variances,
- If the distribution of the groups of the dependent variable generated according to these levels is not suitable for normal distribution, then the Mann-Whitney U test is used.

If the selected parameter has 3 or more levels and at the same time;

- If the distribution of the groups of the dependent variable generated according to these levels is suitable for normal distribution and the variances of the groups are homogeneous, then one-factor analyses of variance (ANOVA),
- If the distribution of the groups of the dependent variable generated according to these levels is suitable for normal distribution but the variances of the groups are not homogeneous, then the Welch or Brown-Forsythe tests,
- If the distribution of the groups of the dependent variable generated according to these levels is not suitable for normal distribution, then the Kruskal-Wallis test is used.

In the second step of the analysis, discriminant analysis was selected to determine the price range of a real estate according to the parameters that affect the dependent variable (real estate sale prices). It should be noted that previous studies mainly used regression models for determining the sale prices of real estates, however; these models provide exact results based on the parameters that are included in the model. Moreover, most of the variables that affect the sale prices are categorical, and, thus, need to be considered as dummy inputs. For example, a real estate can have a city or sea view. This categorical variable cannot be included in the regression model since it does not have a continuous value. Therefore, a model that can include categorical variables is suitable for explaining the sale price of a real estate. On the other hand, time series models only use historical data to construct estimation models. These models rely on the data regarding the sale prices of real estates and do not incorporate any variables in the estimation model.

In discriminant analysis, the independent variables should show a multivariate normal distribution and there should not be a multicollinearity problem between independent variables (Atakan & Karabulut, 2003). In the analysis, functions that allow independent variables and dependent variable groups to be separated from each other are created. Through these functions, the following can be determined:

- The discriminant variables that has the most impact on the separation between groups.
- The variance in the dependent variable caused by the independent variables.
- The group of a new input data.

4. Results

4.1. *Effects of parameters*

In order to determine whether the parameters are effective on real estate sale prices, it is necessary to analyze whether the difference between the averages shown in Table 2 is statistically significant. In this study, the Shapiro-Wilk normality test was performed for the dependent variable to find whether the data is normally distributed. Since the p value was obtained as 0.000 at a significance level of 0.05, it was determined that it is not suitable to normal distribution. In addition, the coefficient of skewness was found to be a positive value (3.238), showing that the distribution is right-skewed. Subsequently, $1/x$ transformation is applied to the data in order to make the data suitable for normal distribution. The result of the Shapiro-Wilk normality test performed for the transformed data was obtained as 0.395 at the significance level of 0.05, in other words, the distribution was suitable to normal distribution. Therefore, the inverse of real estate sale prices is used as the dependent variable.

The results of the Shapiro-Wilk normality tests of the groups of the dependent variable generated according to the parameters of level 2 are normally distributed except for the "property type" parameter. Next, Levene's tests were conducted. The results of these tests are presented in Table 2.

Table 2. Levene's test results for level 2 parameters.

| | Levene's statistic | df1 | df2 | <i>p</i> |
|------------------|--------------------|-----|-----|--------------|
| Property type | 1.086 | 1 | 136 | 0.299 |
| Listed by | 0.353 | 1 | 136 | 0.553 |
| Building status | 0.000 | 1 | 136 | 0.987 |
| Furnished | 0.969 | 1 | 136 | 0.327 |
| Balcony | 2.576 | 1 | 136 | 0.111 |
| View | 1.267 | 1 | 136 | 0.262 |
| Number of bath- | 0.278 | 1 | 136 | 0.599 |
| Air conditioning | 8.310 | 1 | 136 | 0.005 |
| Playground | 8.391 | 1 | 136 | 0.004 |
| Pool | 10.354 | 1 | 136 | 0.002 |
| Gym | 0.462 | 1 | 136 | 0.498 |

The results show the variances between the dependent variable groups are homogeneous except for the "air conditioning", "playground" and "pool" parameters. Subsequently an independent t-test that does not accept the variances as equal, was performed to understand whether the "air conditioning", "playground" and "pool" parameters have an effect on the dependent variable. In addition, a t-test that accepts the variances as equal was performed to determine whether the remaining parameters have an effect. The results are presented in Table 3.

Table 3. T-test results.

| | t | df | <i>p</i> (2-partial) |
|---------------------|--------|-----|----------------------|
| Listed by | -2.206 | 136 | 0.029 |
| Building status | 0.521 | 136 | 0.603 |
| Furnished | -2.732 | 136 | 0.007 |
| Balcony | 1.716 | 136 | 0.088 |
| View | 5.264 | 136 | 0.000 |
| Number of bathrooms | -0.109 | 136 | 0.913 |
| Heating type | -3.772 | 136 | 0.000 |
| Playground | 2.702 | 136 | 0.008 |
| Pool | -3.823 | 136 | 0.000 |
| Gym | -5.779 | 136 | 0.000 |

The results show that "property status", "view", "air conditioning", "playground", "pool" and "gym" parameters have an effect on the dependent variable. In addition, the effect of the "property type" parameter on sale prices was analyzed with the non-parametric Mann-Whitney U test. While the statistical value of this test was found to be 921, the p value was found to be 0.000 at the 0.05 significance level. Since this value is less than 0.05, it can be said that the "property type" parameter is also effective on the dependent variable.

Regarding the parameters "move-in status", "floor number", "building age", "number of rooms", "gross area" and "net area" with a level of 3 or more are normally distributed according to the results of Shapiro-Wilk normality tests. Next, Levene's tests were conducted and the results are presented in Table 4.

Table 4. Levene's test results for parameters of level 3 and more.

| | Levene's statistics | df1 | df2 | <i>p</i> |
|-----------------|---------------------|-----|-----|----------|
| Floor number | 0.461 | 5 | 132 | 0.805 |
| Move-in status | 4.215 | 2 | 135 | 0.017 |
| Building age | 2.795 | 3 | 134 | 0.043 |
| Number of rooms | 0.907 | 2 | 135 | 0.406 |
| Gross area | 1.187 | 2 | 135 | 0.308 |
| Net area | 0.688 | 2 | 135 | 0.504 |

The results show that the variances between the dependent variable groups for the "move-in status" and "building age" parameters are not homogeneous whereas these variances were found to be homogeneous for the remaining parameters. Therefore, Brown-Forsythe was performed in order to understand whether the "move-in status" and "building age" parameters have an effect on the dependent variable whereas ANOVA tests were performed for the remaining parameters. Table 5 and Table 6 present the results of Brown-Forsythe and ANOVA tests, respectively. The results show that "floor number", "room number", "gross area", "net area" and "building age" parameters have an effect on the dependent variable.

Table 5. Results of Brown-Forsythe test.

| | Statistics | df1 | df2 | <i>p</i> |
|----------------|------------|-----|--------|----------|
| Move-in status | 0.882 | 2 | 86.875 | 0.417 |
| Building Age | 9.128 | 3 | 82.110 | 0.000 |

Table 6. Results of ANOVA analysis.

| | Square | df | Mean | F statistics | <i>p</i> |
|-----------------|--------|----|-------|--------------|----------|
| Floor number | 0.000 | 5 | 0.000 | 7.996 | 0.000 |
| Number of Rooms | 0.000 | 2 | 0.000 | 3.867 | 0.023 |
| Gross area | 0.000 | 2 | 0.000 | 7.792 | 0.001 |
| Net area | 0.000 | 2 | 0.000 | 6.808 | 0.002 |

4.2. Discriminant analysis

Discriminant analysis was conducted to classify the sale prices of real estates based on the effective parameters. For this analysis, real estate prices were divided into 3 groups as 50 thousand to 100 thousand dollars; 100 thousand to 150 thousand dollars and 150 thousand to 200 thousand dollars. The distribution of these data by groups show that real estates whose sale prices vary between 50 and 100 thousand dollars constitute the majority (56%) of the data whereas sale prices which vary between 100-150K and 150-200K are 28% and 16%, respectively. Table 7 shows the initial results of the discriminant analysis.

Table 7. The equality test of the group averages.

| | Wilks' lambda | F | df1 | df2 | <i>p</i> |
|------------------|---------------|-------|-----|-----|----------|
| Listed by | 0.964 | 2,50 | 2 | 135 | 0.086 |
| Property type | 0.693 | 29,84 | 2 | 135 | 0.000 |
| Floor number | 0.833 | 13,54 | 2 | 135 | 0.000 |
| Furniture status | 0.889 | 8,44 | 2 | 135 | 0.000 |
| View | 0.782 | 18,84 | 2 | 135 | 0.000 |
| Building age | 0.854 | 11,53 | 2 | 135 | 0.000 |
| Number of rooms | 0.936 | 4,62 | 2 | 135 | 0.011 |
| Gross area | 0.954 | 3,24 | 2 | 135 | 0.042 |
| Net area | 0.996 | 0,28 | 2 | 135 | 0.755 |
| Air conditioning | 0.873 | 9,86 | 2 | 135 | 0.000 |
| Playground | 0.881 | 9,08 | 2 | 135 | 0.000 |
| Gym | 0.837 | 13,18 | 2 | 135 | 0.000 |
| Pool | 0.795 | 17,40 | 2 | 135 | 0.000 |

The results show that at least 2 of the 3 groups (50K-100K, 100K-150K, 150K-200K) of the dependent variable differ according to all the independent variables except for "listed by" and "net area" parameters. The reason for "listed by" and "net area" parameters were not effective on the dependent variable in this analysis is the high correlation with other parameters. Therefore, these two parameters are excluded from discriminant analysis. The test results of the discriminant analysis performed by assuming the same covariance matrix for all groups with the remaining parameters are shown in Table 8.

Table 8. Discriminant analysis test results.

| | |
|---------------|-----------|
| Box's M | 204.370 |
| Approximate F | 1.921 |
| df1 | 91 |
| df2 | 18205.037 |
| <i>p</i> | 0.000 |

Table 8 shows the accuracy of the assumption whether the covariance matrices of the groups are equal to each other. Since the obtained *p* value (0.000) is less than 0.05 significance level, it can be concluded that the covariance matrices of the groups are not equal. Therefore, discriminant analysis was repeated using separate covariance matrices for the groups. The results of the functions obtained according to the discriminant analysis are shown in Table 9 and Table 10.

Table 9. Eigen values.

| Function | Eigen value | Variance (%) | Cumulative (%) | Canonical Correlation |
|----------|-------------|--------------|----------------|-----------------------|
| 1 | 1.659 | 85.6 | 85.6 | 0.790 |
| 2 | 0.278 | 14.4 | 100.0 | 0.467 |

Table 10. Wilks' Lambda results.

| Function | Wilks' Lambda | Chi-square | df | <i>p</i> |
|----------|---------------|------------|----|----------|
| 1 | 0.294 | 159.054 | 22 | 0.000 |
| 2 | 0.782 | 31.913 | 10 | 0.000 |

Wilks' Lambda values in Table 10 show the unexplained variance. In this table, it is seen that the *p* values of both functions are less than 0.05 significance level. Therefore, it can be said that these two functions separate the groups in a statistically significant way. The square of the canonical correlation values in Table 10 shows the explained variance. Accordingly, it can be interpreted that in the dependent variable, the first function can explain 62% variance and the second function can explain 22% variance. Also, the Eigen value in the same table indicates the power of the function to explain the variance in all variables. Accordingly, it is seen from the variance percentages section in the table that 85.6% of the total variance is explained by the first function and 14.4% by the second function. The standardized canonical discriminant function coefficients (weights) of these functions as well as importance order of parameters are presented in Table 11.

Table 11. Weights and importance order of parameters in classification functions.

| Order of im- portance | Function 1 | | Function 2 | |
|--------------------------|------------------|--------|------------------|--------|
| | Parameters | Weight | Parameters | Weight |
| 1 | Property type | 0.958 | Number of rooms | 1.364 |
| 2 | Gross area | 0.846 | Gross area | -0.990 |
| 3 | Pool | -0.620 | Furniture status | 0.518 |
| 4 | Building age | 0.400 | Pool | -0.460 |
| 5 | Playground | 0.391 | Property type | 0.281 |
| 6 | Number of rooms | 0.361 | Air conditioning | 0.265 |
| 7 | Air conditioning | -0.281 | Playground | 0.259 |
| 8 | Floor number | 0.190 | Gym | -0.247 |
| 9 | View | 0.154 | Building Age | 0.075 |
| 10 | Furniture status | 0.121 | Floor Number | 0.064 |
| 11 | Gym | -0.027 | View | 0.055 |

Discriminant functions are found as follows:

$$\text{Discriminant_score_1} = 0.958*\text{property type} + 0.190*\text{floor number} + 0.121*\text{furniture status} + 0.154*\text{view} + 0.400*\text{building age} + 0.361*\text{number of rooms} + 0.846*\text{gross area} - 0.281*\text{air conditioning} - 0.027*\text{gym} + 0.391*\text{playground} - 0.620*\text{pool} \quad (1)$$

$$\text{Discriminant_score_2} = 0.281*\text{property type} + 0.064*\text{floor number} + 0.518*\text{furniture status} + 0.055*\text{view} + 0.075*\text{building age} + 1.364*\text{number of rooms} - 0.990*\text{gross area} + 0.265*\text{air conditioning} - 0.247*\text{gym} + 0.259*\text{playground} - 0.460*\text{pool} \quad (2)$$

Table 11, Equations (1) and (2) show that the order of importance and weight of the parameters in both functions differ. For example, the order of importance of the parameters "room number", "air conditioning", "furniture status" and "gym" is lower in function 1 than function 2. The order of importance of other parameters is higher in function 1. It is seen that the most important parameter in function 1 is the "property type", and in function 2, the order of importance of "property type" is 5. The second important factor in both functions is the gross area parameter.

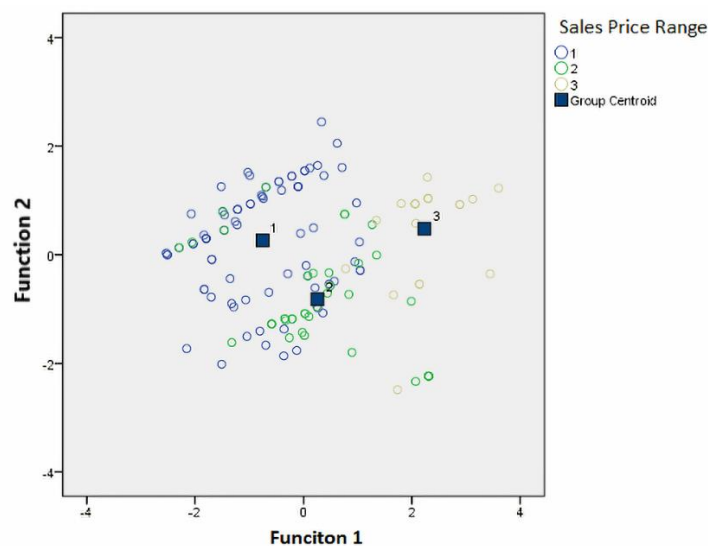


Figure 1. The positions of canonical discriminant functions on the centroid graph.

Figure 1 explains which groups each function can distinguish from each other. Based on function 1, it is seen that the distance between the 50-100 group and the 150-200 group is the farthest. Therefore, it can be said that function 1 separates groups 50-100 and 150-200 from each other. Based on function 2, it is seen that the furthest distance is between 100-150 and 150-200 groups. Therefore, it can be said that function 2 separates 100-150 and 150-200 groups from each other.

Table 12 shows how many of the real estates would be classified in the correct sale price group when placed randomly.

Table 12. Probabilities of random placement in groups.

| Sale price range | Priority probability | Situations in analysis | | Random correct classification probability |
|------------------|----------------------|------------------------|----------|---|
| | | Non-weighted | Weighted | |
| 50-100 | 0.565 | 78 | 78.000 | 55.09 (0,565*78*1,25) |
| 100-150 | 0.275 | 38 | 38.000 | 13.06 (0,275*38*1,25) |
| 150-200 | 0.159 | 22 | 22.000 | 4.37 (0,159*22*1,25) |
| Total | 1.000 | 138 | 138.000 | |

The random correct classification probabilities of real estates are obtained by multiplying the primary probability with the non-weighted situation. In order for the classification to be considered successful, the classification must correctly assign more properties to each group than the number of real estates that can be classified with this random probability. Therefore, a margin of error of 25% is added to this value in the literature (McLachlan, 2004). The random correct classification probabilities with this margin of error are shown in Table 12.

Table 13 shows how many real estates were classified into which groups with the functions obtained as a result of the discriminant analysis.

Table 13. Classification results.

| | | Estimated group memberships | | | Total |
|--------|---------|-----------------------------|---------|---------|-------|
| | | 50-100 | 100-150 | 150-200 | |
| Number | 50-100 | 67 | 11 | 0 | 78 |
| | 100-150 | 13 | 22 | 3 | 38 |
| | 150-200 | 0 | 3 | 19 | 22 |
| % | 50-100 | 85.9 | 14.1 | 0 | 100 |
| | 100-150 | 34.2 | 57.9 | 7.9 | 100 |
| | 150-200 | 0 | 13.6 | 86.4 | 100 |

In Table 14 it is seen that 67 of the 78 real estates in the 50-100 group, 22 of the 38 properties in the 100-150 group and 19 of the 22 properties in the 150-200 group were classified in the correct sale price groups. The real estates classified correctly for each group are more than the real estates that can be classified correctly in a random manner, with addition of the margin of error seen in Table 13. It can be concluded that the classification is successful and the sale price groups of 78.3% of the real estates can be correctly estimated by the functions based on independent variables.

5. Conclusions

It is well-known that determining the exact sale price of a real estate is a challenging task due to the complex and dynamic relationship between parameters that affect the real estate appraisal. Furthermore, intangible assets play an important role on

the sale prices. The fact that intangible assets are appraised subjectively, functions that determine sale price ranges based on effective parameters are required for determining the sale prices of real estates. This study provides discriminant functions that enable to find the correct classification of the sale price groups of real estates. Within this context, a total of 138 real estates, which are located in Bayraklı district of Izmir is taken as a case study. Statistical analyses were conducted to identify the most effective parameters on the sale prices and the discriminant analysis was performed to develop the functions. Main findings of the study are as follows:

1. The property type, from whom it is listed, furniture status, the view, air conditioning, playground, pool and gym, the number of rooms, floor number of the real estate, building age, gross and net areas have a statistically significant effect on the sale price of the real estate.
2. Building status, move-in status, balcony and the number of bathrooms do not have a statistically significant effect on the sale price of the real estate.
3. The sale price of the real estate has a positive correlation with the floor number, the number of rooms, gross and net areas; while it has a negative correlation with the building age.
4. Being a flat in a high-rise apartment, being furnished, having central heating system, pool and gym significantly increase the sale price of the real estate.
5. With discriminant analysis, 78.3% of the real estates are correctly classified in sale price groups between 50-100, 100-150 and 150-200 thousand dollars.
6. The most important parameter in the 1st function, which distinguishes real estates with a sale price of 50-100 thousand dollars from those with a sale price of 150-200 thousand dollars, is the "property type" while the least important parameter is the "gym".
7. The most important parameter in the 2nd function, which separates real estates with a sale price between 100-150 thousand dollars and those between 150-200 thousand dollars, is the "number of rooms", while the least important parameter is the "view".

The findings of this study can be utilized by real estate appraisers or researchers to understand the parameters that are effective in the valuation of multi-storey real estates and to identify the correct sale price group via the classification models. Future studies can focus on investigating the effective parameters and developing models that can classify the sale price groups of different types of real estates such as residential and commercial buildings.

References

- Afşar, A. (2017). Determining The Parameters of Housing Prices Using Hedonic Model: A Case Study in Eskisehir. *The Journal of Institute of Social Sciences*, (37), 195–205, (in Turkish).
- Appraisal Institute. (2015). *Standards of Valuation Practice*. Chicago: Appraisal Institute. www.appraisalinstitute.org/professional-practice/ethics-and-standards/standard-of-professionalappraisal-practice/
- Atakan, C., & Karabulut, İ. (2003). Discrimination Based on Depth. *Selçuk University Journal of Science Faculty*, 22, 53–63, (in Turkish).
- Ayan, E., & Erkin, H. C. (2014). Hedonic Modeling for a Growing Housing Market: Valuation of Apartments in Complexes. *International Journal of Economics and Finance*, 6(3), 188–199. doi:10.5539/ijef.v6n3p188
- Bhagat, N., Mohokar, A., & Mane, S. (2016). House Price Forecasting using Data Mining. *International Journal of Computer Applications*, 152(2), 23–26. doi:10.5120/ijca2016911775
- Bin, J., Tang, S., Liu, Y., Wang, G., Gardiner, B., Liu, Z., & Li, E. (2017). Regression model for appraisal of real estate using recurrent neural network and boosting tree. 2017 2nd IEEE International Conference on Computational Intelligence and Applications, ICCIA 2017, 2017-Janua, 209–213. doi:10.1109/CIAPP.2017.8167209
- Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional science and urban economics*, 22(3), 453-474.
- Cao, K., Diao, M., Wu, B. (2019). A big data-based geographically weighted regression model for public housing prices: a case study in Singapore. *Ann. Am. Assoc. Geograph.* 109 (1), 173–186.
- Chiarazzo, V., Caggiani, L., Marinelli, M., & Ottomanelli, M. (2014). A neural network based model for real estate price estimation considering environmental quality of property location. *Transportation Research Procedia*, 3(July), 810–817. Doi: 10.1016/j.trpro.2014.10.067

- Çiçek, U., & Hatırlı, S. A. (2015). Factors Affecting Housing Prices in Isparta Analysis Using Hedonic Price Model. Mehmet Akif Ersoy University Journal of Social Sciences Institute, 7(13), 98–114, (in Turkish).
- Daşkiran, I. (2015). Estimation of The Hedonic Pricing Model of Factors Affecting Housing Demand in The City of Denizli. The Journal of International Social Research, 8(37), 850–857, (in Turkish).
- Erdem, N. (2017). An Approach for Turkish Real Estate Valuation System. Journal of Geomatics, 2(1), 18–36, (in Turkish).
- Ferlan, N., Bastic, M., & Psunder, I. (2017). Influential factors on the market value of residential properties. Engineering Economics, 28(2), 135–144. doi:10.5755/j01.ee.28.2.13777
- Kangallı Uyar, S. G., & Yayla, N. (2016). The Estimation Of Housing Prices Within The Frame Of Hedonic Pricing Approach By Spatial Econometric Models: The Case Of Istanbul Housing Market. Social Sciences (NWSASOS), 11(4), 326–342. doi:10.1007/s11187-017-9901-7, (in Turkish).
- Koramaz, T. K., & Dokmeci, V. (2012). Spatial Determinants of Housing Price Values in Istanbul. European Planning Studies, 20(7), 1221–1237. doi:10.1080/09654313.2012.673569
- Manjula, R., Jain, S., Srivastava, S., & Rajiv Kher, P. (2017). Real estate value prediction using multivariate regression models. IOP Conference Series: Materials Science and Engineering, 263(4), 1–7. doi:10.1088/1757-899X/263/4/042098
- Marmolejo-Duarte, C. (2017). Does urban centrality influence residential prices? An analysis for the Barcelona metropolitan area. Revista de La Construcción, 16(1), 57–65. doi:10.7764/RDLC.16.1.57
- Marmolejo-Duarte, C., & Chen, A. (2019). The evolution of energy efficiency impact on housing prices. An analysis for Metropolitan Barcelona. Revista de La Construcción, 18(1), 156–166. doi:10.7764/RDLC.18.1.156
- McLachlan, G. (2004). Discriminant Analysis and Statistical Pattern Recognition. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Sahibinden.com. (2019). Retrieved from <https://www.sahibinden.com/>. (last access: 9 September 2020).
- Sarip, A. G., Hafez, M. B., & Nasir Daud, M. (2016). Application of fuzzy regression model for real estate price prediction. Malaysian Journal of Computer Science, 29(1), 15–27. Doi: 10.22452/mjcs.vol29no1.2
- SPSS software. (2017). IBM, SPSS Statistics for Windows, Version 25.0. IBM Corp., Armonk.
- Uğur, L. O., & Leblebici, N. (2019). Investigation of the effects of energy and water saving benefits on property value in LEED certified green buildings. Technical Journal of Turkish Chamber of Civil Engineers, 30(1), 8753–8776. doi:10.18400/tekderg.312932
- Wang, F., Zou, Y., Zhang, H., & Shi, H. (2019). House Price Prediction Approach based on Deep Learning and ARIMA Model. Proceedings of IEEE 7th International Conference on Computer Science and Network Technology, ICCSNT 2019, 303–307. doi:10.1109/ICCSNT47585.2019.8962443
- Yayar, R., & Karaca, S. S. (2014). Determining Factors Effecting Housing Prices with Hedonic Model: A Case of TR83 Region. Ege Academic Review, 14(4), 509–509. doi:10.21121/eab.2014418008



Copyright (c) 2021. Kuru, M., Erdem, O. and Calis, G. This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).