STAT 6519- Regression Models

PREDICTION MODEL OF REAL ESTATE SALES PRICE

Submitted to:

Dr. Zhaozhi Fan

Submitted by:

Md Rakibul Alam	ID No. 201794937
Towhidul Islam	ID No. 201995392
Shajib Kumar Guha	ID No. 201893129

Table of Content

List of Figures

List of Tables

1 Executive Summary

Regression analysis is one of the most widely used techniques for analyzing multifactor data. It has broad area and usefulness from the conceptually logical process of using to express the relationship between a variable of interest and a set of related predictor variables. Computing plays an important role in regression analysis. In our project R language under RStudio IDE has been used. We predict the residential homes sales price in a mid-western city as a function of various characteristics of the home and surrounding property. 522 sets of data along with 12 variables have been used initially to fit the model. Issues such as multicollinearity has been observed. Residual analysis has been applied to the reduced model and tried to follow the pattern to determine whether the transformation is needed or not. In this work, we also checked possible outliers, high leverage points, and influential points using diagonal elements of hat matrix, COOK'S D, DFFITS, DFBETAS and COVRATIO. On the base of these analyses, 169 observations (influential points) have been removed and finally the reduced model was determined that revealed as the most satisfactory model.

2 Introduction

The regression method is frequently used as a guided approach to data modeling. There are several types of regression modeling:

- Simple linear regression
- Multiple linear regression
- Polynomial regression
- Logistic regression
- Generalized linear regression etc.

Linear statistical methods are widely used as part of this modeling process. In the biological, physical, and social sciences, as well as in business and engineering, linear models are useful in both the planning stages of research and analysis of the resulting data.

For our assumed project, the city tax assessor was interested in predicting residential home sales prices in a mid- western city as a function of various characteristics of the home and surrounding property. Data on 522 arms-length transactions were obtained for home sales during the year 2002. Using this dataset we have developed a linear regression model to predict the real estate sales price.

3 Objective

The objective of our project is to develop a model to predict the real estate sales price for given data set. R language under IDE RStudio has been used to code the model, diagnostics and corresponding treatment.

4 Literature Review

4.1 Regression Model

Regression analysis is one of the powerful statistical methods to find the proper relation within a dataset, genrally, between the independent variables (predictors) and a dependent variable (outcome). Among several methods of regression analysis, linear regression is the basic foundation of modeling history and is largely used for many practical applications.

4.2 R and Rstudio

R is one of the programming language developed in 1995 at the University of Auckland as an environment for statistical computing and graphics. This language used for statistical computing while RStudio uses the R language to develop statistical programs.

4.3 ANOVA

The analyst utilizes the ANOVA test results in an F-test to generate additional data that aligns with the proposed regression models. The ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them. The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples.

4.4 Multicollinearity

In multiple regression, two or more independent variables might be correlated with each other. This situation is referred as collinearity. On the other hand, if there is an extreme situation where collinearity can be found among three or more variables even if no pair of variables has a particularly high correlation is called mulicollinearity. In the presence of multicollinearity, the solution of the regression model becomes unstable.

4.5 R^2

The definition of R-squared is fairly straight-forward; it is the percentage of the response variable variation that is explained by a linear model.The R-squared (R^2) ranges from 0 to 1 and represents the proportion of information (i.e. variation) in the data that can be explained by the model. The adjusted R-squared adjusts for the degrees of freedom.

4.6 Model Adequacy check

The major assumptions we considered so far:

1. The relationship between the response y and regressor is linear, at least approximately
2. Error term has zero mean and constant variance
3. Errors are normally distributed
4. Errors are uncorrelated

The assumptions can be checked with residual diagnostics.

4.7 Transformation

The usual approach for dealing with inequality of variance is to apply a suitable transformation. In practice, transformation of the response is generally employed to stabilize variance.

4.8 Outlier, Leverage point and influential point

An **outlier** is a data for which response variable does not satisfy the trend of the rest of the data.

A data point has high **leverage** if it has "extreme" predictor *x* values. With a single predictor, an extreme *x* value is simply one that is particularly high or low. With multiple predictors, extreme x values may be particularly high or low for one or more predictors, or may be "unusual" combinations of predictor values. The hat matrix plays an important role in identifying influential observations. The diagonal elements of hat matrix may be interpreted as the amount of leverage. We traditionally assume that any observation for which the hat diagonal exceeds twice the average (2*p)/n is remote enough from the rest of the diagonal data to considered a leverage point.

A data point is **influential** has a great influence on the results of a regression model in form of R square and adjusted R square. It is therefore important to detect influential observations and to take them into consideration when interpreting the results To measure the influential points COOK'S D, DFFITS, DFBETAS and COVRATIO are used.

Outliers and high leverage points has good chance to be influential, but we generally have to investigate further to determine whether or not they are actually influential.

5 Methodology & Interpretation

Data on 522 arms-length transactions were obtained for home sales during the year 2002. Each line of the data set has an identification number and provides information on 12 other variables. Description of the dataset has been given in Table- 1.

Table-1: Description of Dataset

Variable Number	Variable Name	Description
1	Identification number	1-522
2	Sales price	Sales price of residence (dollars)
3	Finished square feet	Finished area of residence (square feet)
4	Number of bedrooms	Total number of bedrooms in residence
5	Number of bathrooms	Total number of bathrooms in residence
6	Air conditioning	Presence or absence of air conditioning: 1 if yes; 0 otherwise
7	Garage size	Number of cars that garage will hold
8	Pool	Presence or absence of swimming pool: 1 if yes; 0 otherwise
9	Year built	Year property was originally constructed
10	Quality	Index for quality of construction: 1 indicates high quality; 2 indicates medium quality; 3 indicates low quality
11	Style	Qualitative indicator of architectural style 12
12	Lot size	Lot size (square feet)

13	Adjacent to highway	Presence or absence of adjacency to highway:1 if yes; 0 otherwise

5.1 Data Cleaning

From our inspection result, four categorical variables (qualitative variables that take on values which are names or labels) are found: air conditioning, quality, style and adjacent to the highway. These categorical variable need to be converted to indicator (dummy) variable. Conversion results are shown in Table 2.

Table-2: Data format conversion

Variable	Type of Variable	Value					
Price	Num	360000	340000	250000	205500	275500	...
Area	Num	3032	2058	1780	1638	2196	...
#bedroom	Num	4	4	4	4	4	...
#bathroom	Num	4	2	3	2	3	...
Air_conditioning	Factor levels "NO","YES":	2	2	2	2	2	...
Garage_capacity	Num	2	2	2	2	2	...
Pool	Factor levels "NO","YES":	1	1	1	1	1	...
Year	Num	1972	1976	1980	1963	1968	...
Quality	Factor levels "HIGH","LOW","MEDIUM	3	3	3	3	3	...
Style	Factor levels "1","2","3","4",.."10"	1	1	1	1	7	...
Lot_area	Num	22221	22912	21345	17342	21786	...
Adj_to_highway	Factor levels "NO","YES":	1	1	1	1	1	...
AGE	Num	26	22	18	35	30	...

Before using data to develop model we need to clean our data based on summary statistics as shown in Table 3. Summary statistics are performed on numeric data. Summary statistics show that there is no missing value of any observation. However, the minimum number of bathroom and bedroom in one real estate is zero which needs to be looked into. The mean and median values have no strange difference. However, there is slight positive skewness (1.55) in price variable (mean of price > median of price).

Table-3: Summary Statistics

Numeric Variable	Minimum	Maximum	Mean	Median	Std.deviation	Missing values
Price	84000	920000	277894	229900	137923.4	0
Area	980	5032	2261	2061	711.0659	0
#bedroom	0	7	3.471	3	1.014358	0
#bathroom	0	7	2.642	3	1.064169	0
Garage_capacity	0	7	2.1	2	0.6539705	0
Lot_area	4560	86830	24370	22200	11684.08	0
Age	0	113	31.1	32	17.63792	0

5.2 ANOVA for Model 1

Considering 5% significance level, from ANOVA table for Model-1, it has been found that two of the variables i. e. Pool and Adjacent to highway are statistically insignificant (p value greater than 0.05).

Table- 4: Analysis of Variance Table for Model-1

Variables	Degree of freedom	Sum Sq.	Mean Sq.	F value	Probability(>F)
Area	1	6.6555e+12	6.6555e+12	2032.2755	2.2e-16 ***
Bedroom	1	2.7613e+10	2.7613e+10	8.4316	0.00385 **
Bathroom	1	1.4271e+11	1.4271e+11	43.5771	1.041e-10 ***
Air Conditioning	1	3.3417e+10	3.3417e+10	10.2040	0.00149 **
Garage capacity	1	2.0019e+11	2.0019e+11	61.1288	3.179e-14 ***
Pool	1	1.2314e+08	1.2314e+08	0.0376	0.84632
Quality	2	8.5703e+11	4.2852e+11	130.8490	2.2e-16 ***
Style	9	1.3094e+11	1.4548e+10	4.4424	1.287e-05 ***
Lot Area	1	6.2251e+10	6.2251e+10	19.0086	1.580e-05 ***
Adjacent to Highway	1	7.2275e+09	7.2275e+09	2.2069	0.13802
Age	1	1.5320e+11	1.5320e+11	46.7809	2.321e-11 ***
Residual	509	1.6407e+12	3.2749e+09		
Significant. Codes Range: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’					

5.3 Multicollinearity Check for Model-1

For a given predictor, multicollinearity can be understood by the variance inflation factor (or VIF), which measures how much the variance of a regression coefficient is inflated due to multicollinearity in the model.

As a rule of thumb, a VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity (James et al. 2014). A generalized version of the VIF, called the GVIF, exists for testing sets of predictor variables and generalized linear models. From the Table-5 of R output, it has been found that the GVIF values are in suggested limit. Hence, there can be seen no multicollinearity among the regressors of model 1.

Table- 5: Multicollinearity Check for Model-1

Name of Variables	GVIF
Area	4.665851
Bedroom	1.733497
Bathroom	3.204204
Air conditioning	1.407803
Garage capacity	1.669460
Pool	1.093521
Quality	4.034712
Style	3.227237
Lot area	1.192550
Adj to highway	1.032620
AGE	2.092467

5.4 ANOVA for Model-2

After removal of two insignificant variables (Pool and Adjacent to highway) model 1 has been updated and renamed as Model 2. . From ANOVA table for Model 2 it has been found that all variables are statistically significant.

Table- 5: Analysis of Variance Table for Model-2

Project Data	**Degree of freedom**	Sum Sq.	Mean Sq.	F value	Probability(>F)
Area	1	6.6555e+12	6.6555e+12	2017.6288	2.2e-16 ***
Bedroom	1	2.7613e+10	2.7613e+10	8.3708	0.003978**
Bathroom	1	1.4271e+11	1.4271e+11	43.2630	1.202e-10 ***
Air Conditioning	1	3.3417e+10	3.3417e+10	10.1305	0.001549 **
Garage capacity	1	2.0019e+11	2.0019e+11	60.6883	3.856e-14 ***
Quality	2	8.5698e+11	4.2849e+11	129.8976	2.2e-16 ***
Style	9	1.3055e+11	1.4506e+10	4.3976	1.501e-05 ***
Lot Area	1	6.1416e+10	6.1416e+10	18.6183	1.923e-05 ***
Age	1	1.4332e+11	1.4332e+11	43.4471	1.102e-10 ***
Residual	503	1.6592e+12	3.2987e+09		
Significant .codes range: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

5.5 Initial Fit ($R^2$and adjustedR^2)

The summary statistics of model 2 below tells us the value of $R^2$and adjusted R^2 are 0.8326 and 0.8266 respectively (see Appendix-2). Hence, variance of almost 83% dataset values can be explained by model 2.

```
Multiple R-squared:  0.8326,   Adjusted R-squared:  0.8266
```

5.6 Model Adequacy Check for Model-2

The normal probability plots (in RStudio, QQplot considered for normality test) help in verifying the assumption of normal distribution. This Figure-1 shows sharp upward and downward curves at both extremes which does not look ok. This indicates that the distribution is heavy tailed.

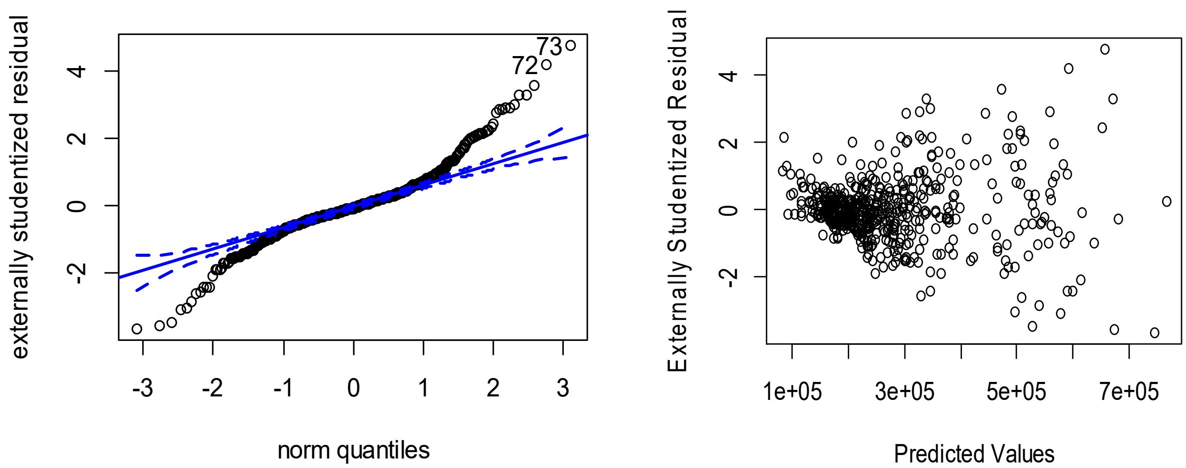

Figure-1: QQ plot for normality test (left) & Residual Plot (right)

A plot of residual versus the corresponding fitted (predicted) values is useful for detecting several common types' model inadequacies. Here, externally studentized residual vs predicted value plot takes the outward-opening funnel pattern as shown in Figure-2. This figure implies that the variance of the errors is not constant and the variance is the expanding with predicted sales price.

5.7 Transformation of Model-2

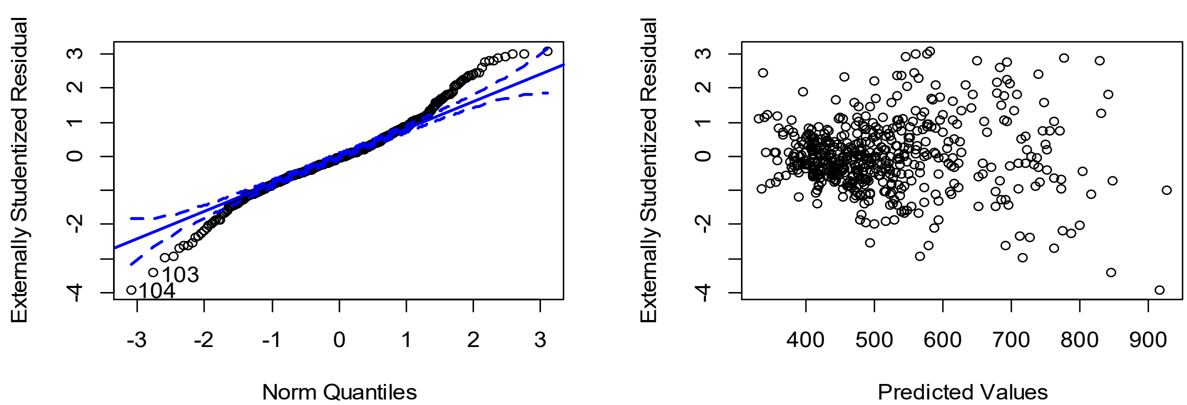

Figure-2: QQ plot for normality test & Residual Plot for Model 2_1

Additionally, Log transformation was implemented on model 2. The transformed model (model 2_2) of model 2 for the real estate sales data with transformed variable y^{*}= log(y) shows much better residual plot (see Figure-3).

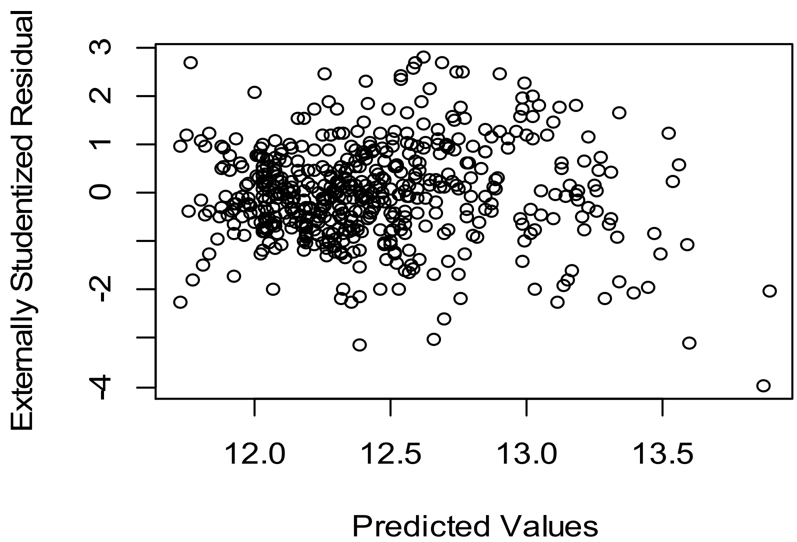

Figure-3: Residual Plot for Model 2_2

Moreover, the summary statistics shows that this transformed model (model 2_2) has satisfactory value of R^2=0.8423 and adjusted R^2 = 0.8367 (see Appendix-4).

```
Multiple R-squared:  0.8423,   Adjusted R-squared:  0.8367
```

ANOVA result shows that regressor '#bedroom' is insignificant. Therefore, new model 3 removing this variable was made up. Unfortunately, model 3 shows higher number of influential points and less improvement of R square and adjusted R square. Therefore, finally transformed model (model 2_2) has been selected for further operations.

5.8 Identification of Leverage points and Influential points

Leverage point: For our data set of project the value of (2*p)/n is 0.04 (where, parameter P=10 and total number of observations, n=522). On the base of these there are 22 points has been selected as potential leverage points. R output column named as "hat" of Appendix 5 indicates the value of diagonal elements of hat matrix.

DFBETAS: For the project data set the value of $2/\sqrt{\text{n}}$ = 0.0875. Appendix 5 reveals the values of DFBETAS which exceed 0.0875.

DFFITS: We have also investigated the deletion influence of the ith observation on the predicted value or fitted value. This leads to DFFITS method. Any observation for which

DFFITS>$2\sqrt{p/n}$ warrants attention. Here, the value of $2\sqrt{p/n}$ =0.277. Table of Appendix-5 reveals the values of DFFITS which exceed 0.277.

A Measure of Model Performance (COVRATIO): The diagnostics COOK'S D, DFFITS, DFBETAS provide insight about the effect of observations on estimated coefficients and fitted values. They do not provide information about overall precision of estimation. If COVERATIO > 1+(3p)/n or if COVERATIO < 1-(3p)/n, them the ith point should be influential. Here, 1+(3p)/n= 1.06 and 1-(3p)/n= 0.94.In our study, the influential points are considered based on COVRATIO method. On the base of COVRATIO 169 observations have been detected. Which has been shown in Appendix-5 and graphically presented in Figure- 4.

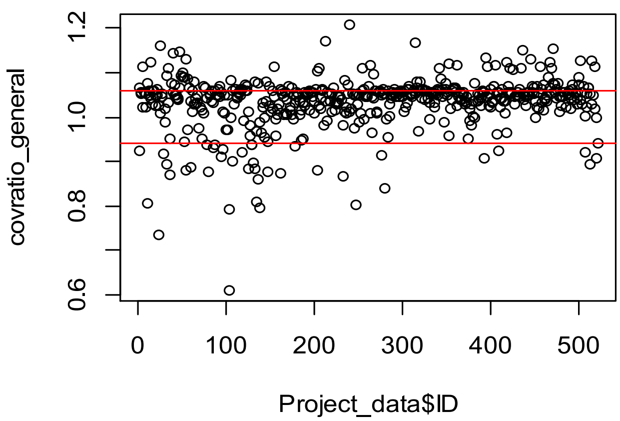

Figure-4: Residual Plot for Model 2_2 (showing COVRATIO limit)

5.9 Removal of Influential Points and Model 2_3

After the removable of 169 observations, data set for model 2_2 has been updated and renamed as model 2_3. The residual plot reveals more satisfactory pattern (see Figure- 5).

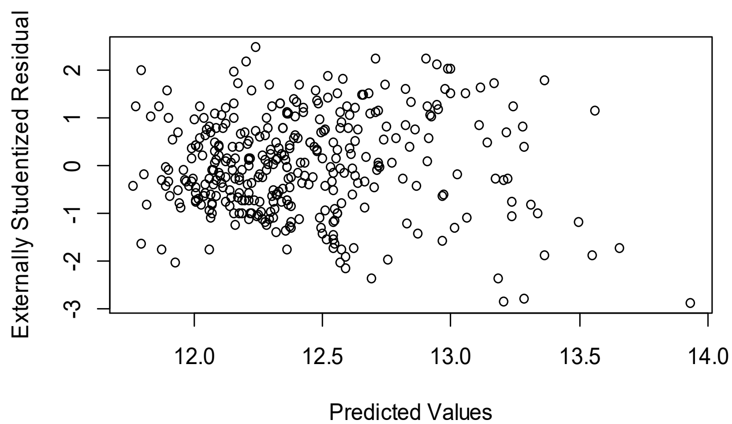

Figure-5: QQ plot for normality test & Residual Plot for Model 2_3

Again on the base of COVRATIO, influential points have been detected (see Figure- 6). But from a practically viewpoint, is fairly small in amount.

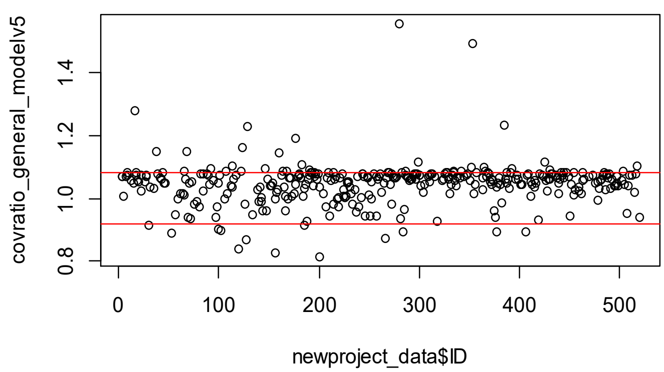

Figure-6: Residual Plot for Model 2_3 (showing COVRATIO limit)

6 Conclusion and Recommendation

After going through all possible processes such as ANOVA test, multicollinearity, normality test, residual analysis, transformation of model, possible outliers detection, high potential leverage points and influential points identification for developing a suitable regression model that can predict the residential homes sales, and the final model has been shown below.

Log (Real Estate Sales Price) = 11.69 + .0003278 Area + .006826 Bedroom + .07356 Bathroom + .01653 Air Condition Yes + .04477 Garage Capacity - .2697 Quality Low - .221 Quality Medium - .05775 Style 2 + .011 Style 3- .2294 Style 4 -.09534 Style 5 -.1155 Style 6 -.09976 Style 7 + .000002 Lot Area – .004229 Age

- The number of bathroom has the highest influence to increase the real estate sales price.
- Low Quality has the highest influence to decrease the real estate sales price.
- The existence of pool and location of real estate adjacent to the highway make negligible effect on the real estate prices.
- Among significant contributors of real estate sales price, lot area has the least impact.
- The model validation part can be performed in future.
- The robust regression model can be implemented to reduce the impact of extreme outliers.
- This model is for Midwestern city, for other cities this model may not work. Therefore, it should be careful to use this model in other cities.

Appendix

1. Dataset

1.	1	360000	3032	4	4	1	2	0	1972	2	1	22221	0
2.	2	340000	2058	4	2	1	2	0	1976	2	1	22912	0
3.	3	250000	1780	4	3	1	2	0	1980	2	1	21345	0
4.	4	205500	1638	4	2	1	2	0	1963	2	1	17342	0
5.	5	275500	2196	4	3	1	2	0	1968	2	7	21786	0
6.	6	248000	1966	4	3	1	5	1	1972	2	1	18902	0
7.	7	229900	2216	3	2	1	2	0	1972	2	7	18639	0
8.	8	150000	1597	2	1	1	1	0	1955	2	1	22112	0
9.	9	195000	1622	3	2	1	2	0	1975	3	1	14321	0
10.	10	160000	1976	3	3	0	1	0	1918	3	1	32358	0
11.	11	190000	2812	7	5	0	2	1	1966	3	7	56639	0
12.	12	559000	2791	3	4	1	3	0	1992	1	1	30595	0
13.	13	535000	3381	5	4	1	3	0	1988	1	7	23172	0
14.	14	525000	3459	5	4	1	2	0	1978	1	5	35351	0
15.	15	299900	2090	3	3	1	2	0	1987	2	1	24025	0
16.	16	527000	3232	5	5	1	2	0	1984	2	6	21445	0
17.	17	169900	1502	2	2	1	2	0	1956	2	1	28958	0
18.	18	335250	2747	3	4	1	2	0	1993	2	7	22241	0
19.	19	323900	2890	4	3	1	2	0	1954	2	7	41992	0
20.	20	200000	1825	3	3	1	2	0	1957	2	1	30266	0
21.	21	211000	1578	4	3	1	2	0	1986	2	2	18829	0
22.	22	212000	1763	3	3	1	2	0	1959	2	1	24726	0
23.	23	245000	2517	4	3	1	2	0	1965	2	1	23261	0
24.	24	140400	1872	3	2	1	2	0	1985	2	3	24017	0
25.	25	295000	3266	3	3	1	2	0	1908	2	6	24881	0
26.	26	170900	2020	1	2	1	1	0	1956	2	1	21385	0
27.	27	229000	2164	4	2	1	2	0	1965	2	1	28291	0
28.	28	218500	2080	3	2	1	2	1	1959	2	1	14752	0
29.	29	160000	2208	2	2	1	2	0	1985	2	7	8058	0
30.	30	259000	3048	6	4	1	3	0	1960	2	7	29307	0
31.	31	164500	1460	3	2	1	2	0	1978	2	1	9999	0
32.	32	280000	2540	3	2	0	2	0	1940	2	5	42428	0
33.	33	154000	2208	2	2	1	2	0	1985	2	7	6746	0
34.	34	272000	2560	4	2	1	3	0	1977	2	5	36100	0
35.	35	180000	2061	4	2	0	2	0	1958	2	1	20138	0
36.	36	157500	1980	3	2	1	2	0	1957	2	1	32519	0
37.	37	242500	3308	5	4	1	2	0	1928	2	5	47323	0
38.	38	182000	2616	5	3	0	2	0	1955	3	5	11123	0
39.	39	178000	1460	4	2	1	2	0	1961	3	1	27095	0
40.	40	171900	1580	2	1	0	1	0	1951	3	4	12417	0
41.	41	165500	1460	3	2	1	2	0	1960	3	1	22493	0
42.	42	183500	1540	3	2	1	2	0	1992	3	3	15801	0
43.	43	135000	1388	2	1	0	2	0	1951	3	1	26106	0
44.	44	175000	1624	3	2	1	2	0	1948	3	1	39219	0
45.	45	149500	1580	2	1	1	2	0	1966	3	1	11166	0
46.	46	177500	1820	3	2	1	2	0	1960	3	1	22104	0
47.	47	155000	1733	4	1	1	1	0	1936	3	4	22398	0
48.	48	145000	1896	3	2	0	2	0	1925	3	6	32753	0
49.	49	178000	2038	2	2	0	2	0	1918	3	7	47884	0
50.	50	156000	1436	3	2	0	3	0	1920	3	1	43594	0
51.	51	159000	1690	3	2	0	1	0	1922	3	5	28518	0
52.	52	160000	1496	2	2	0	1	0	1900	3	5	43335	0
53.	53	112000	1668	2	1	0	1	0	1948	3	1	19612	0
54.	54	84000	980	1	1	0	1	0	1951	3	1	17686	0
55.	55	155000	2562	3	2	0	2	0	1885	3	7	40800	0
56.	56	360000	2304	5	4	1	3	0	1978	2	1	70240	0
57.	57	104000	1268	2	1	0	1	0	1947	3	1	21067	0
58.	58	420000	2283	3	3	1	3	0	1997	1	1	18524	1
59.	59	355000	2060	2	3	1	2	0	1997	2	1	38623	1
60.	60	165000	2087	2	2	1	2	0	1966	2	1	24764	1
61.	61	244000	2081	4	2	1	2	0	1980	2	3	24993	1
62.	62	179900	1696	3	3	1	2	0	1978	2	2	22294	1
63.	63	253000	2222	4	2	0	2	0	1955	2	1	71527	1
64.	64	200000	2110	5	3	1	2	0	1957	2	1	15332	1
65.	65	200000	1774	4	2	0	2	0	1963	3	1	15528	1
66.	66	147700	1592	3	2	1	2	0	1957	3	1	11221	1
67.	67	188700	1748	3	2	1	2	0	1972	3	1	23939	1
68.	68	177000	1985	3	1	0	2	0	1948	3	1	69975	1
69.	69	585000	2558	2	4	1	3	1	1984	1	3	24601	0
70.	70	549900	4000	6	5	1	3	1	1979	1	10	23595	0
71.	71	675000	3942	4	3	1	2	0	1990	1	7	18920	0
72.	72	830000	3889	4	4	1	3	0	1991	1	7	28378	0
73.	73	920000	3857	4	5	1	3	0	1997	1	1	32793	0
74.	74	855000	4756	4	4	1	3	0	1990	1	7	22215	0
75.	75	585500	3302	4	3	1	3	0	1982	1	7	26463	0
76.	76	399000	2629	3	3	1	2	0	1989	1	9	24778	0
77.	77	790000	4418	5	5	1	3	0	1997	1	7	22024	0
78.	78	665000	4746	4	4	1	3	0	1996	1	7	23368	0
79.	79	725000	3242	3	3	1	3	0	1989	1	1	27173	0
80.	80	647000	2464	3	3	1	3	0	1992	1	1	31703	0
81.	81	780000	4419	4	5	1	7	0	1987	1	1	56127	0
82.	82	657500	3877	3	3	1	3	0	1992	1	7	24639	0

262.	262	261000	2404	5	3	1	3	1	1973	2	2	17791	0
263.	263	174500	1840	3	2	1	2	0	1960	2	1	16356	0
264.	264	244900	2107	3	2	0	2	0	1947	2	6	30050	0
265.	265	274900	2788	4	3	1	2	0	1984	2	7	18960	0
266.	266	214000	2416	3	3	1	2	0	1984	2	3	15594	0
267.	267	283000	2430	3	4	1	3	0	1984	2	6	18164	0
268.	268	177900	1584	3	2	1	2	0	1989	2	2	13947	0
269.	269	237500	1873	3	3	1	2	0	1978	2	3	21998	0
270.	270	202150	1644	4	3	1	2	0	1976	2	2	19499	0
271.	271	235000	2400	4	3	1	2	0	1976	2	7	44347	0
272.	272	217000	2094	4	3	1	2	0	1984	2	3	18617	0
273.	273	285000	2561	3	3	1	3	0	1981	2	7	18985	0
274.	274	217500	1752	3	2	1	2	0	1976	2	2	24053	0
275.	275	210000	1738	5	3	1	2	0	1983	2	3	15206	0
276.	276	183340	2068	3	2	1	3	0	1977	2	1	24325	0
277.	277	252000	2428	4	3	1	2	0	1966	2	7	22727	0
278.	278	237000	2090	4	2	1	2	0	1969	2	1	22055	0
279.	279	205000	1820	3	2	1	2	0	1944	2	2	28023	0
280.	280	285000	3219	4	4	1	2	0	1944	2	6	28200	0
281.	281	210000	2654	3	5	1	2	0	1962	2	1	28882	0
282.	282	280000	1802	4	3	1	2	0	1956	2	1	27700	0
283.	283	207000	1765	3	2	1	2	0	1976	2	2	22983	0
284.	284	221000	2786	4	3	1	2	0	1976	2	7	22875	0
285.	285	257000	1794	3	2	1	2	0	1960	2	3	21691	0
286.	286	274000	2768	3	3	0	2	0	1921	2	7	26268	0
287.	287	262000	2288	3	2	1	2	0	1963	2	3	16975	0
288.	288	204400	2028	3	2	1	2	0	1951	2	1	26777	0
289.	289	254900	2620	5	3	1	2	0	1966	2	7	27989	0
290.	290	244000	1644	3	3	1	2	0	1980	2	2	32164	0
291.	291	213000	1888	3	2	1	3	0	1958	2	1	14757	0
292.	292	240000	2116	4	3	1	2	0	1964	2	7	22041	0
293.	293	235000	2313	4	3	1	2	0	1972	2	7	24705	0
294.	294	206000	1824	4	2	1	2	0	1959	2	1	14748	0
295.	295	237000	1942	4	3	1	2	0	1972	2	2	23105	0
296.	296	274000	2184	4	3	1	2	0	1977	2	3	19090	0
297.	297	275000	2578	3	3	1	2	0	1965	2	7	22299	0
298.	298	218400	2036	4	3	1	2	0	1960	2	7	21996	0
299.	299	156000	1384	2	1	0	2	0	1961	2	1	26706	0
300.	300	220000	1826	4	3	1	2	0	1952	2	1	19870	0
301.	301	171500	1681	3	2	1	1	0	1957	2	1	15985	0
302.	302	180000	1726	3	2	1	2	0	1962	2	1	26769	0
303.	303	204000	1910	3	2	1	2	0	1958	2	3	15423	0
304.	304	307000	2664	4	3	1	2	0	1962	2	7	22684	0
305.	305	265000	2116	3	3	1	2	0	1976	2	2	33344	0
306.	306	209900	2030	3	3	1	2	0	1959	2	1	21914	0
307.	307	173000	1940	3	2	0	2	0	1956	2	3	11610	0
308.	308	189000	1676	3	3	0	2	0	1965	2	3	21780	0
309.	309	222500	2120	3	2	1	2	1	1959	2	1	17883	0
310.	310	265000	2152	4	3	1	2	0	1987	2	1	26075	0
311.	311	264670	1984	4	3	1	2	0	1966	2	1	31204	0
312.	312	200750	1575	3	3	1	2	0	1957	2	1	25543	0
313.	313	227900	1798	3	3	1	2	0	1978	2	2	17820	0
314.	314	255000	2017	3	1	0	1	0	1958	2	3	86571	0
315.	315	208500	1904	3	3	1	2	0	1978	2	2	15559	0
316.	316	226900	1718	3	3	1	2	0	1976	2	2	49613	0
317.	317	215000	1776	4	2	0	1	0	1980	2	2	22839	0
318.	318	222950	2609	4	3	1	2	0	1961	2	1	26087	0
319.	319	239900	2226	3	3	1	2	0	1955	2	2	13520	0
320.	320	176000	1556	2	1	1	3	0	1959	2	3	15623	0
321.	321	228000	1764	2	2	1	2	0	1985	2	3	8105	0
322.	322	204900	1626	3	3	1	2	1	1968	2	2	15288	0
323.	323	258000	2012	5	3	1	2	0	1967	2	2	21303	0
324.	324	241850	2090	4	4	1	2	0	1963	2	7	22010	0
325.	325	198500	2192	4	2	1	2	0	1963	2	2	22851	0
326.	326	243000	2228	4	4	1	2	0	1965	2	7	21881	0
327.	327	187000	1825	4	2	1	2	0	1963	2	2	15810	0
328.	328	233000	2132	4	3	1	2	0	1966	2	2	17159	0
329.	329	205000	2160	3	2	1	2	0	1983	2	7	16555	0
330.	330	205000	1974	3	2	1	2	0	1954	2	1	26196	0
331.	331	189000	1696	3	3	1	2	0	1968	2	1	31851	0
332.	332	204900	2132	3	3	1	2	0	1969	2	7	23986	0
333.	333	239000	1814	4	3	1	2	1	1979	2	2	24698	0
334.	334	193000	1796	3	2	1	2	0	1963	2	2	29281	0
335.	335	260000	2268	5	3	1	2	0	1971	2	3	35240	0
336.	336	188000	1719	2	3	1	2	0	1960	2	1	22009	0
337.	337	190500	1704	3	2	1	2	0	1984	2	3	22583	0
338.	338	230000	2142	3	2	1	2	0	1959	2	1	22223	0
339.	339	240000	1705	4	2	1	2	0	1972	2	1	42322	0
340.	340	235000	1752	4	3	1	2	0	1978	2	2	39267	0
341.	341	275000	2554	5	3	1	2	1	1966	2	7	22381	0
342.	342	205000	1650	3	2	1	2	0	1965	2	3	27235	0
343.	343	280000	2816	5	3	1	2	0	1946	2	5	29109	0

83.	83	578000	3808	5	4	1	3	0	1982	1	7	23324	0
84.	84	500000	3376	5	4	1	2	0	1947	1	7	18452	0
85.	85	484530	2940	3	3	1	3	1	1979	1	7	20639	0
86.	86	609000	2654	5	3	1	3	0	1997	1	1	12821	0
87.	87	635000	2672	4	3	1	3	0	1995	1	1	28049	0
88.	88	519000	3386	4	4	1	3	0	1994	1	7	24008	0
89.	89	625100	3648	4	4	1	3	0	1992	1	7	26604	0
90.	90	585444	3455	4	5	1	3	0	1995	1	7	22468	0
91.	91	399900	3321	3	4	1	3	1	1971	1	7	15012	0
92.	92	389900	2817	4	3	1	3	0	1996	1	7	31214	0
93.	93	649000	3210	3	5	1	3	0	1995	1	1	30033	0
94.	94	535000	3588	4	5	1	3	0	1987	1	7	22530	0
95.	95	640000	2705	3	3	1	3	0	1994	1	1	22196	0
96.	96	600000	2344	4	3	1	2	0	1925	1	1	86004	0
97.	97	582500	4264	5	4	1	4	0	1995	1	7	24983	0
98.	98	545000	2615	3	3	1	3	0	1996	1	1	21722	0
99.	99	480000	3608	6	4	1	3	0	1981	1	7	25219	0
100.	100	595000	2479	4	4	1	3	0	1989	1	1	29805	0
101.	101	610000	3251	3	4	1	3	1	1985	1	1	25570	0
102.	102	570000	2547	2	3	1	3	0	1996	1	1	21789	0
103.	103	479000	5032	7	3	1	3	0	1989	1	7	22000	0
104.	104	545000	4973	6	6	1	3	1	1987	1	7	56139	0
105.	105	335000	2582	4	3	1	2	0	1966	1	2	23256	0
106.	106	629000	3139	6	4	1	2	1	1977	1	1	21810	0
107.	107	505500	3516	4	4	1	3	0	1979	1	7	19867	0
108.	108	528750	2129	0	0	1	3	0	1992	1	1	37414	0
109.	109	370000	2936	4	4	1	3	0	1987	1	7	16437	0
110.	110	529000	3351	5	4	1	3	0	1994	1	7	24216	0
111.	111	490000	3136	4	4	1	3	0	1989	1	7	27158	0
112.	112	535000	3525	4	5	1	3	0	1996	1	7	27501	0
113.	113	612000	3917	6	5	1	3	0	1995	1	7	37972	0
114.	114	442500	2702	4	3	1	3	0	1991	1	1	39643	0
115.	115	500000	3644	3	4	1	3	0	1984	1	7	21895	0
116.	116	539000	3072	4	4	1	3	0	1992	1	1	25158	0
117.	117	545500	3233	4	4	1	3	0	1991	1	7	22961	0
118.	118	424000	2918	4	4	1	3	0	1988	1	7	22003	0
119.	119	325000	3266	4	3	1	3	0	1985	1	7	16640	0
120.	120	367000	2940	4	7	1	2	0	1988	1	7	22003	0
121.	121	470000	3430	3	4	1	2	1	1966	1	7	25018	0
122.	122	393000	2472	4	4	1	3	0	1987	1	1	21784	0
123.	123	530000	2878	4	4	1	3	0	1992	1	1	68351	0
124.	124	400000	2537	3	3	1	2	0	1993	1	1	11053	0
125.	125	403500	3858	4	4	1	3	0	1987	1	7	22224	0
126.	126	550000	2742	3	3	1	3	0	1991	1	1	22306	0
127.	127	380000	3460	5	4	1	2	0	1972	1	1	18571	0
128.	128	500000	3836	5	4	1	3	0	1982	1	5	48465	0
129.	129	465000	4453	7	5	1	2	0	1974	1	7	15595	0
130.	130	451500	4080	5	4	1	3	0	1983	1	7	22134	0
131.	131	336000	3301	3	4	1	2	0	1977	1	3	18741	0
132.	132	550000	3828	4	5	1	2	1	1975	1	1	17051	0
133.	133	450000	2973	4	3	1	2	0	1980	2	7	21999	0
134.	134	440000	2821	5	4	1	2	1	1962	2	1	32914	0
135.	135	515000	2950	5	3	1	2	0	1969	2	1	21598	0
136.	136	415000	2362	3	3	1	2	0	1977	2	3	21604	0
137.	137	380000	3092	3	4	1	2	0	1978	2	3	20081	0
138.	138	489500	2866	4	4	0	2	0	1982	2	7	22424	0
139.	139	478000	3369	5	4	1	3	0	1981	2	7	21161	0
140.	140	460000	3068	4	4	1	3	0	1988	2	7	18289	0
141.	141	379900	2380	3	3	1	3	0	1998	2	1	21999	0
142.	142	390000	2225	2	3	1	3	0	1997	2	1	38722	0
143.	143	338000	2655	3	3	1	2	0	1948	2	1	21960	0
144.	144	675000	3855	4	4	1	3	0	1996	2	7	35845	0
145.	145	440000	2892	4	4	1	2	0	1977	2	6	35839	0
146.	146	500000	3832	4	3	1	2	0	1952	2	1	28722	0
147.	147	470000	3164	6	4	1	2	0	1982	2	7	20505	0
148.	148	317500	2620	3	2	1	3	0	1925	2	7	12266	0
149.	149	430000	3076	6	4	1	3	0	1984	2	7	26648	0
150.	150	430000	4022	6	4	1	3	0	1986	2	7	18429	0
151.	151	475000	3377	4	3	1	3	0	1997	2	7	22495	0
152.	152	389000	2858	5	4	1	1	0	1989	2	7	23981	0
153.	153	400000	3540	4	4	1	3	1	1980	2	7	18012	0
154.	154	395000	3045	4	3	1	1	0	1991	2	7	34356	0
155.	155	395000	4150	4	3	0	3	0	1934	2	7	21778	0
156.	156	296000	1778	2	2	1	2	0	1991	2	7	24022	0
157.	157	387500	2412	2	3	1	3	0	1986	2	3	22676	0
158.	158	353000	2668	4	3	1	2	0	1978	2	2	18384	0
159.	159	350000	2274	4	3	1	3	0	1986	2	1	22049	0
160.	160	437632	2936	4	3	1	3	0	1980	2	5	22844	0
161.	161	447500	2526	2	2	1	0	0	1996	2	1	28248	0
162.	162	318500	2449	4	4	1	2	0	1985	2	3	22075	0
163.	163	352000	3131	4	4	1	2	0	1988	2	7	15209	0
164.	164	295000	2536	3	3	1	3	0	1987	2	7	39427	0
165.	165	313500	3314	2	3	1	3	0	1984	2	7	24339	0
166.	166	330000	2230	3	2	1	2	0	1986	2	1	24798	0
167.	167	400000	2631	4	4	0	2	0	1985	2	7	44885	0
168.	168	325000	2638	4	3	1	2	0	1978	2	3	25747	0
169.	169	340000	2756	4	3	1	2	1	1973	2	7	22546	0
170.	170	399900	3262	5	4	1	2	0	1978	2	7	25541	0
171.	171	249900	1936	4	4	1	3	0	1987	2	3	12850	0
172.	172	389000	3148	4	4	1	3	1	1969	2	7	16587	0
173.	173	364500	2616	3	4	1	2	0	1977	2	7	32655	0
174.	174	357500	3630	4	3	1	2	0	1969	2	7	23283	0
175.	175	295000	1954	3	3	1	3	1	1962	2	3	19300	0

344.	344	190000	1919	3	4	1	2	0	1938	2	7	20093	0
345.	345	232500	2080	3	2	1	2	0	1968	2	1	32021	0
346.	346	259500	2108	4	4	1	2	0	1978	2	7	24685	0
347.	347	275000	2480	3	3	1	2	0	1964	2	1	22144	0
348.	348	183900	1746	3	2	1	2	0	1974	2	2	52136	0
349.	349	290000	2703	3	3	0	4	0	1963	2	1	43599	0
350.	350	217950	1640	4	2	1	2	0	1979	2	3	21314	0
351.	351	220000	2196	4	3	1	2	0	1972	2	7	17899	0
352.	352	185000	1701	3	2	1	1	0	1982	2	2	21938	0
353.	353	288000	2250	3	2	1	2	0	1949	2	4	23684	0
354.	354	197500	2502	4	2	1	2	0	1964	2	6	23749	0
355.	355	179975	1762	4	3	1	2	0	1959	2	1	15742	0
356.	356	195000	2016	4	3	1	2	0	1963	2	2	18102	0
357.	357	228400	1904	3	3	1	2	1	1976	2	2	14945	0
358.	358	194750	1652	5	2	1	2	0	1960	2	1	24644	0
359.	359	195000	2042	4	3	0	2	0	1963	2	2	21849	0
360.	360	210000	2019	4	3	1	3	0	1960	2	2	14837	0
361.	361	239550	2791	5	3	1	3	0	1946	2	4	22863	0
362.	362	242000	2514	4	3	1	1	0	1953	2	5	17535	0
363.	363	185000	1746	3	2	1	2	0	1984	2	3	12386	0
364.	364	175000	1930	3	2	1	2	0	1956	2	2	15923	0
365.	365	165000	1552	3	1	1	3	0	1959	2	3	27068	0
366.	366	185000	1566	4	2	1	2	0	1993	2	2	13504	0
367.	367	173194	1669	3	2	1	2	0	1964	2	2	24643	0
368.	368	205150	1814	3	3	1	2	0	1978	2	2	18714	0
369.	369	214200	1794	3	3	1	3	0	1976	2	2	24308	0
370.	370	182500	1691	3	2	1	2	1	1968	2	2	21961	0
371.	371	205000	1834	4	3	1	2	0	1959	2	1	30726	0
372.	372	208000	1984	5	3	1	2	0	1961	2	1	22047	0
373.	373	225000	1966	4	3	1	2	0	1962	2	1	24871	0
374.	374	170000	1669	3	2	1	2	0	1967	2	2	21253	0
375.	375	216000	2132	4	3	1	2	0	1976	2	1	41332	0
376.	376	180000	2007	4	3	1	2	0	1959	2	3	15992	0
377.	377	169200	1964	4	2	1	2	0	1964	2	7	18162	0
378.	378	213000	2325	4	3	1	3	0	1973	2	3	16699	0
379.	379	210000	2196	4	3	1	2	0	1965	2	7	29329	0
380.	380	185000	2061	3	2	1	2	0	1956	2	2	25379	0
381.	381	179900	1828	3	2	1	2	0	1956	2	3	37150	0
382.	382	196000	1956	4	3	1	2	0	1968	2	1	20486	0
383.	383	219900	1852	6	3	1	2	0	1968	2	2	20800	0
384.	384	159900	1795	1	2	1	2	0	1980	2	11	26467	0
385.	385	170000	1580	2	1	0	1	0	1950	3	5	10799	0
386.	386	169900	1708	3	1	0	1	0	1950	3	1	11413	0
387.	387	189500	1700	4	2	0	2	0	1953	3	1	14023	0
388.	388	195000	1742	1	1	1	2	0	1961	3	1	18250	0
389.	389	215000	1890	4	2	1	2	0	1961	3	1	22110	0
390.	390	171000	1512	2	1	0	1	0	1956	3	1	14774	0
391.	391	179900	1840	3	1	0	2	0	1953	3	1	40832	0
392.	392	120000	1060	2	1	0	2	0	1947	3	1	15001	0
393.	393	175000	1540	3	2	0	2	0	1957	3	1	45458	0
394.	394	232900	1550	4	2	1	2	1	1962	3	2	14998	0
395.	395	229900	2787	4	2	1	1	0	1922	3	5	39558	0
396.	396	174900	1528	2	2	1	2	0	1982	3	1	25193	0
397.	397	168900	1928	2	2	0	2	0	1941	3	5	26393	0
398.	398	229500	2329	3	2	1	2	0	1960	3	7	28179	0
399.	399	236000	1940	4	3	1	2	0	1959	3	1	15073	0
400.	400	205500	2114	5	2	1	2	0	1966	3	7	14526	0
401.	401	212000	1799	3	2	1	2	0	1962	3	2	16210	0
402.	402	205000	1864	3	2	1	2	0	1940	3	1	25628	0
403.	403	193000	1581	3	2	1	2	0	1956	3	1	15064	0
404.	404	180000	1652	3	2	1	2	0	1959	3	3	21875	0
405.	405	184000	1592	4	2	0	2	0	1977	3	1	25943	0
406.	406	144900	1520	3	1	0	1	0	1953	3	4	36359	0
407.	407	255000	1792	2	2	1	2	0	1955	3	1	31257	0
408.	408	137000	1464	2	1	0	1	0	1957	3	1	14999	0
409.	409	178000	1702	3	2	0	2	0	1961	3	1	21898	0
410.	410	296000	2180	3	2	1	2	0	1952	3	1	29617	0
411.	411	186500	1486	2	2	1	2	0	1958	3	1	18479	0
412.	412	170000	1364	2	1	0	2	0	1942	3	1	26369	0
413.	413	219000	1540	4	2	1	2	0	1977	3	2	30691	0
414.	414	188000	1608	4	2	1	2	0	1969	3	2	19380	0
415.	415	195250	1668	3	1	1	2	0	1956	3	3	17060	0
416.	416	175000	1944	3	2	1	2	0	1951	3	1	43562	0
417.	417	215000	1883	4	2	1	2	1	1956	3	1	19932	0
418.	418	197500	2215	4	1	0	2	0	1948	3	6	25540	0
419.	419	249900	1916	2	2	1	2	0	1954	3	1	20576	0
420.	420	180000	1508	2	2	1	2	0	1959	3	1	32469	0
421.	421	174900	1809	3	1	1	2	0	1958	3	2	16782	0
422.	422	189900	1958	4	3	0	2	0	1935	3	5	22788	0
423.	423	154000	1592	2	1	1	2	0	1951	3	1	10332	0
424.	424	150000	1636	2	1	1	2	0	1950	3	1	10000	0
425.	425	189900	1800	3	2	0	2	0	1964	3	2	13566	0
426.	426	157000	1600	3	1	0	2	0	1950	3	5	10807	0
427.	427	182000	1550	3	1	1	2	0	1966	3	1	15100	0
428.	428	187650	1578	3	1	1	2	0	1958	3	1	14631	0
429.	429	175000	1644	3	1	1	2	0	1956	3	1	12999	0
430.	430	189900	1556	3	2	1	1	0	1959	3	3	19840	0
431.	431	175000	1672	3	1	1	2	0	1949	3	1	22617	0
432.	432	159900	1650	2	2	1	1	0	1957	3	1	14997	0
433.	433	184900	1676	3	2	0	2	0	1956	3	1	16156	0
434.	434	174900	1960	2	2	0	2	0	1947	3	1	16953	0
435.	435	143000	1649	3	1	1	1	0	1951	3	5	20096	0
436.	436	164900	1728	3	1	1	1	0	1950	3	4	10999	0

176.	176	274500	1926	5	3	1	2	0	1986	2	7	26418	0
177.	177	259000	2556	3	2	1	2	0	1957	2	1	80886	0
178.	178	415000	2282	5	4	1	2	0	1987	2	3	23003	0
179.	179	443000	3314	3	4	1	3	0	1986	2	7	22012	0
180.	180	249000	2001	3	3	1	2	0	1981	2	3	23812	0
181.	181	330000	2607	5	2	1	2	0	1976	2	3	23139	0
182.	182	291000	2840	4	4	1	2	0	1965	2	7	23079	0
183.	183	418000	3036	3	5	0	2	0	1977	2	7	33746	0
184.	184	320000	2240	4	2	1	3	0	1974	2	3	18682	0
185.	185	264000	1788	3	3	0	2	0	1969	2	1	18484	0
186.	186	381000	2620	5	4	1	2	0	1965	2	7	28093	0
187.	187	250000	1480	3	3	1	2	0	1984	2	3	14230	0
188.	188	360000	2588	3	3	1	2	1	1968	2	2	19004	0
189.	189	369500	3138	4	4	1	2	0	1969	2	7	18190	0
190.	190	285400	2460	5	4	1	2	0	1979	2	7	27492	0
191.	191	409000	3566	4	4	1	2	0	1976	2	7	18044	0
192.	192	333000	2692	5	4	1	3	0	1984	2	7	22020	0
193.	193	362000	2958	5	4	1	3	0	1987	2	7	45200	0
194.	194	387500	3164	4	4	1	2	1	1966	2	7	23856	0
195.	195	239000	2058	3	3	1	2	0	1969	2	2	21046	0
196.	196	299900	2717	3	4	1	2	0	1983	2	7	22083	0
197.	197	335000	2920	4	3	1	3	0	1987	2	7	22434	0
198.	198	275000	2554	5	3	1	2	0	1960	2	7	21820	0
199.	199	328000	2805	3	4	1	2	0	1988	2	7	22582	0
200.	200	333000	2736	4	3	1	2	1	1979	2	7	29591	0
201.	201	397000	3516	5	7	1	3	0	1996	2	7	34795	0
202.	202	374800	3536	6	4	1	2	1	1978	2	7	19997	0
203.	203	520000	2138	5	3	1	3	0	1956	2	1	86830	0
204.	204	325000	2718	4	4	1	3	0	1978	2	6	22842	0
205.	205	295000	2178	5	3	1	2	0	1958	2	1	25891	0
206.	206	415000	3152	5	4	1	2	0	1980	2	6	24446	0
207.	207	224900	2611	3	3	1	2	0	1987	2	7	6924	0
208.	208	265000	2060	4	3	1	2	0	1981	2	2	13091	0
209.	209	299900	2448	4	4	1	3	0	1979	2	7	26790	0
210.	210	390000	4050	6	5	1	2	1	1966	2	7	18262	0
211.	211	271000	2414	3	3	1	2	0	1914	2	7	24357	0
212.	212	330000	3072	4	3	1	2	0	1966	2	7	16431	0
213.	213	350000	2525	3	3	1	3	0	1983	2	4	27138	0
214.	214	310000	2866	5	3	1	3	0	1961	2	7	25249	0
215.	215	340000	3246	4	4	1	3	0	1987	2	7	52218	0
216.	216	307000	2707	4	4	1	2	0	1992	2	7	22094	0
217.	217	304000	2300	5	3	1	2	0	1961	2	3	35824	0
218.	218	275900	1860	3	3	1	2	0	1957	2	1	40741	0
219.	219	315000	3636	5	3	1	2	0	1976	2	7	19776	0
220.	220	295000	1910	4	3	1	2	0	1968	2	1	30996	0
221.	221	251010	2280	4	2	1	2	0	1956	2	1	25543	0
222.	222	335000	3386	4	5	0	2	0	1965	2	7	38428	0
223.	223	343500	2324	5	4	1	3	0	1967	2	3	22435	0
224.	224	297000	1970	4	3	1	2	0	1972	2	3	25814	0
225.	225	281000	2062	3	3	1	2	1	1977	2	2	23608	0
226.	226	235000	2617	4	3	1	2	0	1985	2	7	8903	0
227.	227	237000	2612	4	3	1	2	0	1985	2	7	10144	0
228.	228	274900	2472	3	3	1	2	0	1969	2	7	22451	0
229.	229	229900	1922	4	2	1	2	0	1957	2	1	15791	0
230.	230	259000	1852	3	2	1	2	0	1984	2	3	22204	0
231.	231	245000	2239	3	3	1	2	0	1986	2	3	22216	0
232.	232	208000	2068	3	2	1	2	0	1979	2	1	34773	0
233.	233	421000	2101	3	3	0	2	0	1956	2	1	65499	0
234.	234	320000	2200	4	3	1	2	0	1968	2	1	31450	0
235.	235	256000	1972	3	3	0	2	0	1989	2	3	32027	0
236.	236	275000	2007	3	3	1	2	1	1959	2	3	21311	0
237.	237	222000	2612	4	3	1	2	0	1985	2	7	8229	0
238.	238	249900	2124	4	1	1	3	0	1974	2	2	21834	0
239.	239	273500	2612	4	3	1	2	0	1985	2	7	8924	0
240.	240	218500	2548	2	2	1	2	0	1984	2	3	10210	0
241.	241	377000	2767	3	3	1	3	0	1941	2	4	75232	0
242.	242	220000	2025	3	3	1	2	0	1989	2	3	19618	0
243.	243	192900	1956	4	2	1	2	0	1962	2	2	21779	0
244.	244	298750	2460	4	3	1	2	0	1967	2	3	23907	0
245.	245	315000	2764	3	3	1	3	0	1972	2	7	23947	0
246.	246	315000	2004	4	3	1	2	0	1967	2	3	24453	0
247.	247	465900	2852	4	3	1	2	1	1961	2	7	34040	0
248.	248	239500	2096	3	3	1	2	0	1977	2	1	15237	0
249.	249	276000	2330	3	3	1	2	0	1989	2	7	17433	0
250.	250	226000	2520	4	3	1	2	0	1978	2	7	12145	0
251.	251	235000	2528	3	3	1	2	0	1977	2	3	26469	0
252.	252	247000	2030	4	3	1	3	0	1988	2	3	23202	0
253.	253	182000	2208	2	2	1	2	0	1985	2	7	6734	0
254.	254	180000	1500	2	1	1	0	0	1938	2	3	39776	0
255.	255	249000	2078	5	3	1	2	0	1965	2	1	21512	0
256.	256	260000	2442	4	3	1	2	0	1969	2	2	21149	0
257.	257	219900	2612	3	3	1	2	0	1986	2	7	11288	0
258.	258	295000	2268	4	3	1	3	0	1972	2	2	23976	0
259.	259	290000	2734	4	3	1	2	0	1971	2	7	23488	0
260.	260	300000	2228	4	3	1	2	0	1982	2	3	21232	0
261.	261	354900	3000	5	4	1	3	0	1973	2	7	21643	0

437.	437	173500	1586	3	2	1	1	0	1958	3	1	15862	0
438.	438	161800	1592	2	1	0	1	0	1951	3	1	18686	0
439.	439	148000	1514	2	2	1	1	0	1964	3	1	16209	0
440.	440	177000	1952	4	2	1	2	0	1963	3	1	24377	0
441.	441	149900	1550	3	1	1	2	0	1956	3	1	14311	0
442.	442	170000	1544	2	2	0	1	0	1957	3	1	14942	0
443.	443	142000	1566	3	1	0	1	0	1959	3	1	15228	0
444.	444	186900	1650	3	2	1	2	0	1961	3	1	22000	0
445.	445	152900	1392	2	2	0	2	0	1951	3	1	29199	0
446.	446	350000	2981	5	4	1	2	0	1950	3	6	49756	0
447.	447	130000	1412	2	1	0	1	0	1940	3	1	16752	0
448.	448	167900	2180	4	2	0	2	0	1948	3	3	15001	0
449.	449	184900	1704	2	2	1	2	0	1954	3	1	16759	0
450.	450	178000	1600	3	2	1	2	0	1957	3	3	15090	0
451.	451	111000	1276	3	2	0	1	0	1951	3	1	11554	0
452.	452	207000	1666	4	2	0	1	0	1954	3	1	39523	0
453.	453	190000	1760	3	2	1	2	0	1974	3	2	20193	0
454.	454	230000	1836	3	2	0	2	0	1946	3	1	46339	0
455.	455	165000	1636	3	1	0	2	0	1953	3	1	20125	0
456.	456	210000	1748	3	2	1	2	0	1957	3	1	14512	0
457.	457	226000	2556	4	2	0	2	0	1923	3	6	36276	0
458.	458	149900	1511	4	1	1	1	0	1954	3	1	14821	0
459.	459	155000	1524	3	2	1	2	0	1958	3	1	21875	0
460.	460	219900	1821	4	2	1	2	0	1956	3	1	16696	0
461.	461	132000	1596	1	1	0	0	0	1940	3	1	28357	0
462.	462	195000	2392	4	2	0	2	0	1960	3	1	30265	0
463.	463	155900	1748	2	1	1	2	0	1956	3	1	16231	0
464.	464	119900	1384	2	1	0	0	0	1949	3	1	30002	0
465.	465	175000	1628	3	1	1	2	1	1957	3	1	17069	0
466.	466	304000	1911	4	2	1	2	0	1953	3	1	86248	0
467.	467	190000	1624	4	3	0	1	0	1959	3	3	15002	0
468.	468	229100	1956	3	2	1	2	1	1984	3	6	14710	0
469.	469	187500	2012	4	2	0	2	0	1953	3	5	14925	0
470.	470	173000	1590	2	2	1	2	0	1956	3	1	22336	0
471.	471	170000	1687	2	1	1	1	0	1941	3	4	20925	0
472.	472	179000	1816	4	1	1	2	0	1956	3	3	16508	0
473.	473	162500	1622	3	1	1	1	0	1956	3	3	16120	0
474.	474	172000	1604	4	2	1	2	0	1954	3	1	14964	0
475.	475	153200	1592	4	2	0	1	0	1956	3	1	16396	0
476.	476	220000	1922	3	2	1	2	0	1952	3	1	33579	0
477.	477	174000	1892	3	1	1	1	0	1955	3	3	14712	0
478.	478	200000	1628	3	1	1	2	0	1959	3	1	15412	0
479.	479	161000	1644	2	2	1	2	0	1956	3	1	17030	0
480.	480	135000	1450	3	1	1	2	0	1955	3	1	15868	0
481.	481	190000	1592	3	2	1	2	0	1959	3	1	22748	0
482.	482	153800	1654	2	1	0	2	0	1962	3	1	25874	0
483.	483	144900	1388	2	1	1	2	0	1950	3	1	10568	0
484.	484	165000	1670	3	2	0	2	0	1953	3	1	18525	0
485.	485	186000	1953	3	2	1	1	0	1956	3	1	17129	0
486.	486	167000	2008	3	1	1	2	0	1963	3	1	21860	0
487.	487	189600	1650	3	2	1	2	0	1960	3	1	34724	0
488.	488	170000	1578	3	1	1	2	0	1975	3	1	22485	0
489.	489	189500	1618	3	2	1	2	0	1962	3	1	27539	0
490.	490	153500	1642	3	1	0	2	0	1959	3	1	14901	0
491.	491	159900	2008	1	2	1	2	0	1941	3	1	31657	0
492.	492	158000	1604	3	2	1	1	0	1960	3	1	23534	0
493.	493	175000	2035	4	2	1	3	0	1962	3	7	20131	0
494.	494	147000	1534	3	2	0	1	0	1955	3	3	15361	0
495.	495	155000	1624	2	2	1	2	0	1955	3	1	16721	0
496.	496	150000	1700	4	2	1	1	0	1954	3	1	15391	0
497.	497	165000	1630	3	2	1	2	0	1978	3	2	24963	0
498.	498	147000	1526	3	2	1	1	0	1957	3	1	15007	0
499.	499	146250	1672	3	1	1	2	0	1949	3	1	22617	0
500.	500	177500	1588	4	2	1	2	0	1980	3	2	21925	0
501.	501	153650	1752	4	2	1	2	0	1950	3	6	9126	0
502.	502	199500	1674	4	2	0	2	0	1947	3	1	33237	0
503.	503	186000	1980	3	1	0	1	0	1927	3	6	47679	0
504.	504	139900	1396	1	1	0	2	0	1950	3	1	25879	0
505.	505	160000	1178	1	1	0	2	0	1959	3	1	9941	0
506.	506	125000	1263	2	2	0	0	0	1955	3	1	12357	0
507.	507	359900	2377	5	2	1	2	0	1937	3	6	51005	0
508.	508	184500	1304	3	1	1	2	0	1951	3	1	21305	0
509.	509	155000	1340	3	1	0	0	0	1952	3	1	5666	0
510.	510	150000	1559	2	1	0	2	0	1952	3	1	23999	0
511.	511	146000	1412	1	2	1	0	0	1920	3	1	4560	0
512.	512	129000	1198	2	2	1	2	0	1925	3	1	20918	0
513.	513	145000	1424	2	1	1	2	0	1947	3	1	16414	0
514.	514	200000	1370	4	1	0	1	0	1925	3	4	8000	0
515.	515	149900	1584	3	2	1	2	0	1957	3	1	13514	0
516.	516	132000	1567	2	1	1	1	0	1934	3	4	12249	0
517.	517	136900	1409	2	1	0	1	0	1951	3	1	28421	0
518.	518	137000	1655	2	1	0	1	0	1935	3	1	54651	0
519.	519	185000	1944	3	2	1	2	0	1939	3	6	17999	0
520.	520	133500	1922	3	1	0	2	0	1950	3	1	14805	0
521.	521	124000	1480	3	2	1	2	0	1953	3	1	28351	0
522.	522	95500	1184	2	1	0	1	0	1951	3	1	14786	0

2. Summarty Table of R output for model-2

```
Call:
```

```
lm(formula = Project_data$Price ~ Project_data$Area + Project_data$`#bedroom`
+ Project_data$`#bathroom` + Project_data$Air_conditioning +Project_data$Gara
ge_capacity + Project_data$Quality + Project_data$Style + Project_data$Lot_ar
ea + Project_data$AGE)

Residuals:
    Min      1Q  Median      3Q     Max
-200847  -26187   -3558   22509  262697

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                       1.613e+05  2.487e+04   6.485 2.12e-10 ***
Project_data$Area                 1.007e+02  7.627e+00  13.205  < 2e-16 ***
Project_data$`#bedroom`          -4.613e+03  3.265e+03  -1.413  0.15836
Project_data$`#bathroom`          1.087e+04  4.213e+03   2.580  0.01018 *
Project_data$Air_conditioningYES  3.231e+03  7.954e+03   0.406  0.68470
Project_data$Garage_capacity      9.210e+03  4.970e+03   1.853  0.06446 .
Project_data$QualityLOW          -1.428e+05  1.424e+04 -10.032  < 2e-16 ***
Project_data$QualityMEDIUM       -1.327e+05  1.051e+04 -12.626  < 2e-16 ***
Project_data$Style2              -2.492e+04  9.138e+03  -2.727  0.00661 **
Project_data$Style3              -1.318e+04  8.699e+03  -1.516  0.13026
Project_data$Style4               1.548e+04  1.810e+04   0.855  0.39285
Project_data$Style5              -2.500e+04  1.487e+04  -1.681  0.09331 .
Project_data$Style6              -5.425e+03  1.492e+04  -0.364  0.71637
Project_data$Style7              -4.200e+04  8.585e+03  -4.892 1.34e-06 ***
Project_data$Style9              -8.719e+04  5.819e+04  -1.498  0.13466
Project_data$Style10             -7.644e+04  5.873e+04  -1.302  0.19366
Project_data$Style11             -9.799e+04  5.816e+04  -1.685  0.09263 .
Project_data$Lot_area             1.286e+00  2.339e-01   5.498 6.13e-08 ***
Project_data$AGE                 -1.351e+03  2.049e+02  -6.591 1.10e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 57430 on 503 degrees of freedom
F-statistic:   139 on 18 and 503 DF,  p-value: < 2.2e-16
```

3. Summarry Table of R output for transformed model (model 2_1) of model 2

```
Call:
lm(formula = ystar ~ Project_data$Area + Project_data$`#bedroom` +
    Project_data$`#bathroom` + Project_data$Air_conditioning +Project_data$Ga
rage_capacity + Project_data$Quality + Project_data$Style +Project_data$Lot_a
rea + Project_data$AGE)
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-177.802  -25.567   -2.051   25.277  139.888 

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)                       3.827e+02  2.060e+01  18.574  < 2e-16 ***
Project_data$Area                 8.383e-02  6.317e-03  13.271  < 2e-16 ***
Project_data$`#bedroom`          -1.683e+00  2.704e+00  -0.622  0.53398    
Project_data$`#bathroom`          1.151e+01  3.489e+00   3.297  0.00104 ** 
Project_data$Air_conditioningYES  7.935e+00  6.588e+00   1.205  0.22895    
Project_data$Garage_capacity      8.454e+00  4.117e+00   2.054  0.04053 *  
Project_data$QualityLOW          -1.142e+02  1.179e+01  -9.682  < 2e-16 ***
Project_data$QualityMEDIUM       -9.548e+01  8.704e+00 -10.970  < 2e-16 ***
Project_data$Style2              -2.062e+01  7.569e+00  -2.725  0.00666 ** 
Project_data$Style3              -7.604e+00  7.205e+00  -1.055  0.29178    
Project_data$Style4               1.689e+01  1.499e+01   1.127  0.26034    
Project_data$Style5              -1.561e+01  1.231e+01  -1.267  0.20559    
Project_data$Style6               2.239e+00  1.236e+01   0.181  0.85629    
Project_data$Style7              -3.081e+01  7.111e+00  -4.333 1.78e-05 ***
Project_data$Style9              -4.518e+01  4.820e+01  -0.937  0.34903    
Project_data$Style10             -6.355e+01  4.864e+01  -1.306  0.19202    
Project_data$Style11             -9.498e+01  4.817e+01  -1.972  0.04916 *  
Project_data$Lot_area             1.209e-03  1.937e-04   6.243 9.13e-10 ***
Project_data$AGE                 -1.165e+00  1.697e-01  -6.867 1.94e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 47.57 on 503 degrees of freedom
Multiple R-squared:  0.8449,   Adjusted R-squared:  0.8394 
F-statistic: 152.2 on 18 and 503 DF, p-value: < 2.2e-16.
```

4. Summarry table of R output for transformed model (model 2_2) of model 2

```
Call:
lm(formula = ystar ~ Project_data$Area + Project_data$`#bedroom` + Project_da
ta$`#bathroom` + Project_data$Air_conditioning +Project_data$Garage_capacity
+ Project_data$Quality + Project_data$Style +Project_data$Lot_area + Project_
data$AGE)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.66269 -0.10683 -0.00283  0.10561  0.47938 
```

```
Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)                     1.184e+01  7.556e-02 156.674  < 2e-16 ***
Project_data$Area               2.902e-04  2.317e-05  12.528  < 2e-16 ***
Project_data$`#bedroom`         3.027e-03  9.918e-03   0.305 0.760379    
Project_data$`#bathroom`        4.840e-02  1.280e-02   3.782 0.000174 ***
Project_data$Air_conditioningYES  5.143e-02  2.416e-02   2.129 0.033743 *  
Project_data$Garage_capacity    3.515e-02  1.510e-02   2.328 0.020309 *  
Project_data$QualityLOW        -3.726e-01  4.324e-02  -8.616  < 2e-16 ***
Project_data$QualityMEDIUM     -2.692e-01  3.192e-02  -8.434 3.55e-16 ***
Project_data$Style2            -6.618e-02  2.776e-02  -2.384 0.017488 *  
Project_data$Style3            -1.131e-02  2.642e-02  -0.428 0.668782    
Project_data$Style4             7.231e-02  5.497e-02   1.315 0.188949    
Project_data$Style5            -3.307e-02  4.516e-02  -0.732 0.464355    
Project_data$Style6             3.185e-02  4.533e-02   0.703 0.482536    
Project_data$Style7            -9.339e-02  2.608e-02  -3.581 0.000375 ***
Project_data$Style9            -5.901e-02  1.768e-01  -0.334 0.738611    
Project_data$Style10           -2.301e-01  1.784e-01  -1.290 0.197729    
Project_data$Style11           -3.783e-01  1.766e-01  -2.142 0.032713 *  
Project_data$Lot_area           4.703e-06  7.104e-07   6.620 9.21e-11 ***
Project_data$AGE               -4.150e-03  6.224e-04  -6.669 6.82e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1745 on 503 degrees of freedom
Multiple R-squared:  0.8423,   Adjusted R-squared:  0.8367 
F-statistic: 149.3 on 18 and 503 DF,  p-value: < 2.2e-16
```

Analysis of Variance Table of transformed model (model 2_2) of model 2

```
Response: ystar
                              Df  Sum Sq Mean Sq   F value    Pr(>F)    
Project_data$Area              1 5137953 5137953 2270.4912 < 2.2e-16 ***
Project_data$`#bedroom`        1    2419    2419    1.0692    0.3016    
Project_data$`#bathroom`       1  155977  155977   68.9274 9.494e-16 ***
Project_data$Air_conditioning  1   46461   46461   20.5312 7.337e-06 ***
Project_data$Garage_capacity   1  137666  137666   60.8353 3.607e-14 ***
Project_data$Quality           2  475374  237687  105.0352 < 2.2e-16 ***
Project_data$Style             9   81867    9096    4.0197 5.545e-05 ***
Project_data$Lot_area          1   57040   57040   25.2064 7.163e-07 ***
Project_data$AGE               1  106710  106710   47.1557 1.940e-11 ***
Residuals                    503 1138252    2263                        
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

5. Identification of Leverage points and Influential Points

	dfb.1_	dfb.Pr_$A	dfb.Prjct_dt$`#bd`	dfb.Prjct_dt$`#bt`	dfb.P_$A_	dfb.P_$G
11	0.24	0.18	-0.41	-0.33	0.26	0.06
14	-0.02	0.00	-0.01	0.00	-0.01	0.02
24	-0.11	0.02	0.02	0.12	-0.02	0.03
25	-0.01	0.02	-0.01	0.00	0.01	0.00
36	-0.03	-0.01	0.01	0.07	-0.05	0.01
37	0.24	-0.12	-0.01	-0.09	-0.15	0.04
40	0.04	0.00	-0.02	0.00	-0.06	-0.02
47	0.00	0.01	-0.04	0.02	-0.03	0.02
54	-0.22	0.08	0.14	-0.03	0.16	0.07
55	0.04	-0.01	0.01	0.00	0.01	-0.02
70	0.00	0.00	0.00	0.00	0.00	0.00
76	0.00	0.00	0.00	0.00	0.00	0.00
80	0.20	-0.11	-0.02	-0.02	0.00	0.05
81	0.43	-0.19	0.14	-0.06	0.07	-0.68
96	0.10	-0.23	0.10	0.00	0.20	-0.14
103	0.19	-0.57	-0.44	0.61	0.06	0.07
104	0.40	-0.41	-0.05	-0.17	-0.04	0.18
108	0.39	0.03	-0.29	-0.44	0.05	0.09
120	-0.15	0.28	0.10	-0.49	-0.01	0.14
125	-0.03	-0.08	0.04	0.04	0.00	-0.01
133	-0.01	0.08	0.00	-0.08	0.02	-0.06
135	-0.14	0.24	0.17	-0.09	0.00	-0.09
136	-0.04	0.09	-0.10	0.04	0.02	-0.04
138	0.08	0.00	-0.01	0.11	-0.38	-0.04
148	-0.12	0.03	-0.07	-0.13	0.07	0.27
161	0.21	0.33	-0.20	-0.14	0.07	-0.57
203	-0.17	-0.14	0.19	-0.04	0.10	0.12
213	0.00	0.00	0.00	0.00	0.00	0.01
233	0.05	-0.03	-0.06	0.07	-0.23	-0.03
241	0.02	-0.01	0.01	-0.01	-0.01	-0.01
247	-0.05	0.05	0.00	-0.07	0.07	-0.06
264	0.05	-0.02	0.00	-0.02	-0.06	0.01
281	0.15	-0.07	0.23	-0.42	-0.05	0.07
314	-0.01	-0.01	0.00	0.02	0.01	0.02
353	0.00	0.01	-0.02	-0.01	0.01	0.01
361	0.15	-0.03	-0.10	-0.01	0.01	-0.15
362	0.00	0.00	0.00	0.00	0.00	0.00
384	0.00	0.00	0.00	0.00	0.00	0.00
395	-0.02	0.02	0.00	-0.01	0.03	-0.02
397	0.00	0.00	0.01	0.00	0.00	0.00

406	-0.07	0.03	-0.03	0.02		0.09	0.04
418	0.00	0.00	0.00	-0.01		-0.01	0.00
436	0.00	0.00	0.00	0.00		-0.01	0.01

	dfb.P_$QL	dfb.P_$QM	dfb.P_$S2	dfb.P_$S3	dfb.P_$S4	dfb.P_$S5	dfb.P_$S6	dfb.P_$S7
11	-0.30	-0.08	-0.06	-0.08	-0.03	0.08	0.08	-0.18
14	0.02	0.03	0.00	0.00	0.00	-0.07	0.00	0.01
24	0.03	-0.01	0.03	-0.30	-0.01	-0.04	-0.04	-0.03
25	-0.01	0.00	0.00	0.00	-0.01	-0.01	0.03	-0.01
36	0.07	-0.02	0.10	0.09	0.05	0.04	0.04	0.07
37	-0.05	-0.09	0.01	0.01	0.03	-0.42	0.09	0.12
40	0.00	-0.01	0.01	0.00	0.20	0.00	0.01	0.01
47	0.01	0.02	0.00	0.00	-0.15	0.00	0.00	-0.01
54	0.06	0.07	0.03	0.04	0.03	0.02	-0.01	-0.04
55	0.00	0.01	-0.03	-0.02	0.01	0.02	0.01	-0.04
70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
80	-0.22	-0.24	-0.07	-0.07	0.00	0.01	0.02	-0.05
81	-0.20	-0.13	-0.01	0.01	0.00	0.07	0.12	0.25
96	-0.46	-0.49	0.00	0.04	-0.07	-0.08	-0.03	0.08
103	0.03	0.07	-0.05	-0.01	0.08	0.15	0.12	0.12
104	-0.18	0.04	-0.14	-0.09	0.01	0.14	0.16	0.11
108	-0.33	-0.28	-0.03	-0.05	-0.02	0.06	0.10	0.07
120	0.13	0.25	-0.01	-0.02	-0.06	-0.03	-0.04	-0.12
125	0.09	0.14	-0.04	-0.03	0.00	0.01	0.00	-0.07
133	0.06	0.09	-0.02	-0.02	-0.01	0.00	-0.01	0.08
135	0.08	0.17	-0.14	-0.16	-0.07	-0.12	-0.14	-0.30
136	0.06	0.07	0.00	0.23	0.00	-0.01	-0.02	-0.06
138	0.06	0.13	-0.02	-0.04	0.02	-0.03	-0.01	0.08
148	-0.11	-0.01	0.02	0.03	-0.06	-0.04	-0.03	0.19
161	0.06	0.15	-0.11	-0.17	-0.05	-0.04	-0.05	-0.23
203	0.00	0.03	-0.09	-0.05	-0.03	-0.05	-0.06	-0.01
213	0.01	0.01	0.00	0.00	0.03	0.00	0.00	0.00
233	-0.05	0.05	-0.06	-0.07	-0.02	-0.08	-0.06	-0.07
241	-0.01	-0.01	0.00	0.00	-0.06	0.00	0.00	0.00
247	0.01	0.05	0.00	0.00	-0.03	-0.02	-0.02	0.16
264	-0.05	-0.02	-0.01	-0.01	0.00	0.00	0.13	0.01
281	-0.14	-0.17	0.09	0.12	0.03	0.08	0.13	0.26
314	0.01	0.00	0.00	-0.02	0.00	0.00	0.00	-0.01
353	-0.02	0.01	-0.01	-0.02	0.26	0.00	0.00	-0.02
361	-0.02	-0.07	0.04	0.03	-0.51	0.03	0.05	0.10
362	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00
384	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
395	0.01	0.00	0.01	0.01	-0.01	0.05	-0.01	0.00
397	0.00	0.00	0.00	0.00	0.00	-0.02	0.00	0.00
406	0.01	0.02	-0.01	-0.01	-0.33	0.00	-0.02	-0.03

418	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
436	0.00	0.00	0.00	0.00	-0.04	0.00	0.00	0.00

	dfb.P_$S9	dfb.P_$S10	dfb.P_$S11	dfb.P_$L	dfb.P_$AG	dffit	cov.r	cook.d	hat
11	-0.04	0.03	-0.08	-0.36	0.13	-0.92_*	0.81_*	0.04	0.09
14	0.01	0.01	0.00	-0.01	0.01	-0.09	1.12_*	0.00	0.08
24	0.01	-0.02	0.02	-0.07	0.15	-0.45	0.74_*	0.01	0.02
25	0.00	0.00	0.00	-0.01	0.03	0.06	1.16_*	0.00	0.11
36	0.01	0.00	0.02	-0.05	-0.06	-0.25	0.87_*	0.00	0.01
37	-0.01	0.03	-0.01	-0.09	-0.21	-0.65_*	0.95	0.02	0.08
40	0.00	0.00	0.00	-0.03	-0.03	0.22	1.14_*	0.00	0.11
47	0.00	0.00	0.00	0.00	-0.01	-0.17	1.15_*	0.00	0.10
54	0.02	-0.02	0.02	0.03	0.03	-0.35	0.88_*	0.01	0.02
55	-0.01	0.00	-0.01	-0.01	-0.09	-0.12	1.13_*	0.00	0.09
70	0.00	0.00	0.00	0.00	0.00	NaN	NaN	NaN	1.00_*
76	0.00	0.00	0.00	0.00	0.00	NaN	NaN	NaN	1.00_*
80	-0.05	-0.02	-0.02	0.05	-0.08	0.37	0.88_*	0.01	0.03
81	-0.01	0.07	0.01	-0.09	-0.09	-0.86_*	1.05	0.04	0.15_*
96	-0.05	-0.03	0.01	0.55	0.39	0.88_*	0.93	0.04	0.11_*
103	0.03	0.11	-0.06	0.10	0.05	-0.96_*	0.79_*	0.05	0.09
104	0.03	0.12	-0.04	-0.39	0.01	-0.95_*	0.61_*	0.05	0.05
108	-0.07	0.04	-0.07	0.20	-0.17	0.78_*	0.90	0.03	0.09
120	0.03	0.02	0.00	0.02	-0.04	-0.60_*	0.99	0.02	0.09
125	0.02	0.03	0.00	0.03	-0.01	-0.31	0.88_*	0.01	0.02
133	0.00	0.01	-0.01	0.01	-0.07	0.24	0.88_*	0.00	0.01
135	-0.01	-0.06	0.00	-0.10	0.00	0.43	0.81_*	0.01	0.03
136	0.00	0.00	-0.01	-0.01	-0.02	0.33	0.86_*	0.01	0.02
138	0.01	0.00	0.00	-0.06	-0.14	0.49	0.80_*	0.01	0.03
148	0.03	0.01	0.00	-0.19	0.45	0.60_*	0.88_*	0.02	0.06
161	-0.05	0.00	-0.07	0.14	-0.37	0.75_*	0.87_*	0.03	0.08
203	0.01	0.00	0.00	0.57	0.03	0.70_*	0.88_*	0.03	0.07
213	0.00	0.00	0.00	0.00	-0.01	0.03	1.17_*	0.00	0.11_*
233	-0.01	0.00	-0.02	0.32	-0.01	0.50	0.87_*	0.01	0.04
241	0.00	0.00	0.00	-0.04	-0.01	-0.08	1.21_*	0.00	0.14_*
247	0.01	0.01	0.00	0.11	0.07	0.31	0.80_*	0.01	0.01
264	0.00	0.00	0.00	-0.01	0.01	0.16	1.12_*	0.00	0.08
281	0.01	0.04	0.03	0.04	-0.11	-0.54	0.84_*	0.02	0.04
314	0.00	0.00	0.00	-0.04	0.01	-0.06	1.17_*	0.00	0.11_*
353	0.00	0.00	-0.01	-0.02	0.02	0.28	1.12_*	0.00	0.10
361	-0.01	0.03	-0.01	0.09	-0.10	-0.60_*	1.06	0.02	0.11_*
362	0.00	0.00	0.00	0.00	0.00	-0.01	1.12_*	0.00	0.07
384	0.00	0.00	0.00	0.00	0.00	NaN	NaN	NaN	1.00_*
395	0.00	0.00	0.00	0.01	0.01	0.07	1.13_*	0.00	0.09
397	0.00	0.00	0.00	0.00	0.00	-0.02	1.11_*	0.00	0.07
406	0.00	-0.01	0.00	-0.06	0.08	-0.36	1.12_*	0.01	0.11
418	0.00	0.00	0.00	0.00	0.00	0.02	1.12_*	0.00	0.08

```
436  0.00      0.00       0.00       0.00      0.00      -0.04     1.15_*  0.00   0.10
```

6. Code Used in R

```
#import and attach the data
library(readxl)
Project_data <- read_excel("D:/MUN/STAT 6519/term project/Project_data.xlsx")
attach(Project_data)

#data type identification
str(Project_data)

#summary statistics of data for data cleaning purpose
summary(Project_data)

#conversion to indicator variable
Project_data$Air_conditioning<-as.factor(Project_data$Air_conditioning)
Project_data$Pool<-as.factor(Project_data$Pool)
Project_data$Quality<-as.factor(Project_data$Quality)
Project_data$Style<-as.factor(Project_data$Style)
Project_data$Adj_to_highway<-as.factor(Project_data$Adj_to_highway)

#initial model development
modelv1<-lm(Project_data$Price~Project_data$Area+
        Project_data$`#bedroom`+
        Project_data$`#bathroom`+
        Project_data$Air_conditioning+
        Project_data$Garage_capacity+
        Project_data$Pool+
        Project_data$Quality+
```

```
            Project_data$Style+
            Project_data$Lot_area+
            Project_data$Adj_to_highway+
            Project_data$AGE)

#check multicollinearity
library(car)
vif(modelv1)

#ANOVA
anova(modelv1)

#revised model
modelv2<-lm(Project_data$Price~Project_data$Area+
            Project_data$`#bedroom`+
            Project_data$`#bathroom`+
            Project_data$Air_conditioning+
            Project_data$Garage_capacity+
            Project_data$Quality+
            Project_data$Style+
            Project_data$Lot_area+
            Project_data$AGE)

#ANOVA and summary
anova(modelv2)
summary(modelv2)

#model adequacy check
```

```
prd_modelv2<-modelv2$fitted.values
resid_modelv2<-rstudent(modelv2)
library(car)
qqPlot(resid_modelv2,xlab = 'Norm Quantiles',
    ylab = 'Externally Studentized Residual',
    grid = FALSE)

plot(prd_modelv2,resid_modelv2,xlab = 'Predicted Values',
   ylab = 'Externally Studentized Residual')

#transformation
ystar<-log(Project_data$Price)
model 2_2<-lm(ystar~Project_data$Area+
       Project_data$`#bedroom`+
       Project_data$`#bathroom`+
       Project_data$Air_conditioning+
       Project_data$Garage_capacity+
       Project_data$Quality+
       Project_data$Style+
       Project_data$Lot_area+
       Project_data$AGE)
anova(model 2_2)
summary(model 2_2)

#after tranformation again check model adequacy
prd_model 2_2<-model 2_2$fitted.values
resid_model 2_2<-rstudent(model 2_2)
library(car)
```

```
qqPlot(resid_model 2_2,xlab = 'Norm Quantiles',
       ylab = 'Externally Studentized Residual',
       grid = FALSE)

plot(prd_modelv 2_2,resid_modelv3,xlab = 'Predicted Values',
     ylab = 'Externally Studentized Residual')

#removing insignificant variable and check again
Model 3<-lm(ystar~Project_data$Area+
             Project_data$`#bathroom`+
             Project_data$Air_conditioning+
             Project_data$Garage_capacity+
             Project_data$Quality+
             Project_data$Style+
             Project_data$Lot_area+
             Project_data$AGE)
anova(model 3)
summary(model 3)

prd_model 3_1<-model 3$fitted.values
resid_model 3_1<-rstudent(model 3)
library(car)
qqPlot(resid_model 3,xlab = 'Norm Quantiles',
       ylab = 'Externally Studentized Residual',
       grid = FALSE)

plot(prd_modelv 3,resid_modelv4,xlab = 'Predicted Values',
     ylab = 'Externally Studentized Residual')
```

```
summary(influence.measures(model 3))

#comparison between model 2_2 and model 3
anova(model 2_2,model 3)

#findout influential point
covratio_general<-covratio(model 2_2)
covratio_offlimit<-covratio_general>1.06 | covratio_general<0.94
covratio_offlimit

#removeal influential point
newproject_data<-Project_data[!covratio_offlimit,]
plot(Project_data$ID,covratio_general)+
  abline(h=1.06,col='red')+
  abline(h=.94,col='red')

#final model after removal influential point
newproject_data<-newproject_data[-c(39,44,265),]
ystar_2<-log(newproject_data$Price)
model 2_3<-lm(ystar_2~newproject_data$Area+
              newproject_data$`#bedroom`+
              newproject_data$`#bathroom`+
              newproject_data$Air_conditioning+
              newproject_data$Garage_capacity+
              newproject_data$Quality+
              newproject_data$Style+
              newproject_data$Lot_area+
```

```
            newproject_data$AGE)
anova(model 2_3)
summary(model 2_3)

#check model adequacy for final model
prd_model 2_3<-model 2_3$fitted.values
resid_model2_3<-rstudent(model 2_3)
library(car)
qqPlot(resid_model 2_3xlab = 'Norm Quantiles',
      ylab = 'Externally Studentized Residual',
      grid = FALSE)

plot(prd_model 2_3,resid_model 2_3,xlab = 'Predicted Values',
    ylab = 'Externally Studentized Residual')

covratio_general_model 2_3<-covratio(model 2_3)
covratio_offlimit_model 2_3<-covratio_general>1.08 | covratio_general<0.92
plot(newproject_data$ID,covratio_general_model 2_3)+
  abline(h=1.08,col='red')+
  abline(h=.92,col='red')
```