**Assignment 2\_SDM**

1. **Create a table of relevant predictors, hypothesized direction of effect (+/-), and rationale for each hypothesized effect.** 
   1. **Price Sold – (pricesold)**

The variables which are essential for predicting sold price are as follows -

|  |  |  |
| --- | --- | --- |
| **Predictor** | **Effect** | **Rationale** |
| Age | - | Age of house is a major factor while deciding and selling of price. Cost of house decreases as age increases. Age variable is derived from variable year built(yrblt). |
| Area(lppersqft) | + | Price of house increase as area of house increases. |
| Full Bath | + | Price of house increases as we add additional bathrooms with both WC, Bathtub and Shower. |
| Half Bath | + | Price of house increases as we add powder room to house. |
| Garage | + | Price of house increases with additional garage. |
| Roof | + | Roof has high importance in the state of Florida due to its weather and that’s why it adds more to the total cost. |
| Sale condition(splsale) | - | Price of houses with special conditions like bank owned or short sale usually have less value as there is a probability that these houses are not in good condition. |
| Bedrooms | + | Value of house increases as we add more bedrooms to it. |
| Pool | + | Inclusion of pool, either private or community adds up more value to house. |

* 1. **Number of days required for agent to sale house(adom)**

The variables which are essential for predicting number of days required to sale are as follows –

|  |  |  |
| --- | --- | --- |
| **Predictor** | **Effect** | **Rationale** |
| List Price | + | List price of house is major factor for predicting no. of days required to sell house. Increase in list price increases adom. |
| Age | + | As house becomes old, it becomes more difficult for agent to sell it. |
| Area(lppersqft) | + | Bigger houses are more difficult to sell in less time span. |
| Sale condition(splsale) | + | Disputed houses take more time to sell. |
| Beds | + | Houses with more bedroom increases the time required to sell as cost associated with house increases significantly because of it. |
| Full Bathrooms | - | Houses with more bathrooms sell quickly. |
| Half Bathrooms | - | Houses with powder room sell relatively fast than those without bathrooms. |
| Garages | - | Houses with more garages tend to sell quickly. |
| Pool | - | People desire pool in their house or community in Florida which helps to sell house quickly. |

1. **Run three reasonable models for each DV. Present each model and summarize their output in a compact manner using stargazer.**

**a. For dependent variable Price Sold**

1) m1 = lm(pricesold~age+sqft+lppersqft+bathshalf+bathsfull+garages+roof+splsale+beds + pool)

-Baseline model.

2) m2 = lm(pricesold~age+I(age^2)+sqft+I(sqft^2)+bathsfull+bathshalf+garages+roof+splsale+beds + pool)

-To understand the impact of age and sqft transformations for curve flattening.

3) m3 = lm(pricesold~age+sqft+bathsfull+I(bathsfull^2)+bathshalf+I(bathshalf^2)+garages+I(garages^2)+roof+splsale+beds + pool)

-To understand the impact of bathsfull, bathshalf and garages transformations for curve flattening.

==========================================================================================================

Dependent variable:

-----------------------------------------------------------------------------------

pricesold

(1) (2) (3)

----------------------------------------------------------------------------------------------------------

age -1,257.199\*\*\* (413.332) 11,830.590 (16,215.640) -3,841.081\*\*\* (1,124.426)

I(age2) -255.146 (298.330)

sqft 143.021\*\*\* (2.400) 9.288 (15.932) 136.235\*\*\* (6.847)

lppersqft 3,047.617\*\*\* (54.166)

I(sqft2) 0.018\*\*\* (0.002)

bathshalf 7,544.112\*\*\* (2,381.717) 15,590.160\*\* (6,149.252) -55,290.660\*\*\* (17,615.260)

I(bathshalf2) 54,993.820\*\*\* (14,206.030)

bathsfull 4,766.012\* (2,808.486) 24,596.170\*\*\* (7,183.470) -76,624.340\*\*\* (27,576.280)

I(bathsfull2) 16,914.100\*\*\* (4,482.763)

garages -3,173.809 (2,428.664) 14,602.990\*\* (6,251.984) 60,130.060 (48,792.750)

I(garages2) -9,703.240 (9,367.721)

roofConcrete -36,316.550 (29,540.460) 134,806.900\* (75,848.200) 119,386.900 (79,734.850)

roofOther -23,561.170 (29,433.370) 38,941.460 (75,849.260) 19,632.690 (79,722.060)

roofShake 2,753.411 (29,771.450) 62,366.290 (76,717.350) 45,805.930 (80,658.470)

roofShingle -19,721.850 (20,843.840) 20,184.000 (53,721.520) 12,239.780 (56,463.020)

roofSlate -18,010.820 (25,774.620) 153.431 (66,443.110) -2,972.501 (69,851.400)

roofTile -26,939.390 (20,958.440) 51,242.020 (53,982.730) 37,178.000 (56,741.600)

splsaleBank Owned/REO 13,220.010 (13,124.290) 21,678.120 (34,373.630) 23,330.710 (35,767.040)

splsaleNone 845.489 (12,166.080) 85,086.480\*\*\* (31,733.100) 86,015.790\*\*\* (33,022.600)

splsaleShort Sale 27,286.980\* (14,252.200) 17,752.480 (37,275.940) 21,351.890 (38,749.720)

beds -5,548.293\*\*\* (1,965.004) -8,293.171 (5,288.126) -19,199.060\*\*\* (5,397.855)

poolNone -1,169.349 (3,583.724) -501.390 (9,240.371) 167.953 (9,714.326)

poolPrivate -11,404.690\*\*\* (3,474.890) 28,525.570\*\*\* (9,106.920) 18,229.580\* (9,519.358)

poolPrivate, Community -15,254.550\*\*\* (3,861.966) 30,629.490\*\*\* (10,041.620) 21,101.600\*\* (10,527.890)

Constant -357,011.400\*\*\* (27,982.510) -156,653.200 (219,138.100) 32,561.590 (84,942.640)

----------------------------------------------------------------------------------------------------------

Observations 478 478 478

R2 0.982 0.883 0.872

Adjusted R2 0.982 0.878 0.866

Residual Std. Error 20,614.160 (df = 458) 53,139.590 (df = 457) 55,849.310 (df = 456)

F Statistic 1,347.503\*\*\* (df = 19; 458) 173.237\*\*\* (df = 20; 457) 147.353\*\*\* (df = 21; 456)

==========================================================================================================

**b. For dependent variable days required for agent to sell house -**

1) m11 = lm(adom\_agentdaysonmarket~listprice+age+sqft+splsale + pool + beds + bathsfull + bathshalf + garages)

-Baseline Model.

2) m12 = lm(adom\_agentdaysonmarket~listprice+I(listprice^2)+age+sqft+splsale + pool+ beds + bathsfull + bathshalf + garages)

-To understand the impact of listprice transformations for flattening of curve.

3) m13 = lm(adom\_agentdaysonmarket~listprice+age+I(age^2) +sqft+splsale + pool+ beds + bathsfull + bathshalf + garages)

-To understand the impact of age transformations for flattening of curve.

===============================================================================================

Dependent variable:

------------------------------------------------------------------------

adom\_agentdaysonmarket

(1) (2) (3)

-----------------------------------------------------------------------------------------------

listprice -0.00004 (0.0001) -0.001\*\*\* (0.0001) -0.0001 (0.0001)

I(listprice2) 0.000\*\*\* (0.000)

age -3.371\*\* (1.479) -1.708 (1.413) -44.983\*\* (19.249)

I(age2) 0.772\*\* (0.356)

sqft 0.069\*\*\* (0.012) 0.078\*\*\* (0.012) 0.067\*\*\* (0.012)

splsaleBank Owned/REO 52.223 (46.272) 39.369 (43.726) 70.068 (46.818)

splsaleNone 68.936 (43.045) 80.311\*\* (40.674) 88.075\*\* (43.774)

splsaleShort Sale 116.761\*\* (49.999) 93.915\*\* (47.309) 134.962\*\*\* (50.503)

poolNone 9.764 (12.632) 8.568 (11.929) 9.185 (12.585)

poolPrivate -11.289 (12.058) 10.456 (11.742) -9.567 (12.036)

poolPrivate, Community -16.401 (13.363) 8.119 (13.027) -14.208 (13.349)

beds -11.831\* (7.031) -4.466 (6.710) -9.418 (7.091)

bathsfull -9.336 (9.844) -2.861 (9.335) -10.257 (9.814)

bathshalf -13.613 (8.382) -12.343 (7.917) -14.127\* (8.352)

garages -15.965\* (8.529) -5.715 (8.167) -14.399\* (8.526)

Constant 37.331 (63.143) 28.121 (59.636) 577.783\*\* (257.078)

-----------------------------------------------------------------------------------------------

Observations 478 478 478

R2 0.197 0.286 0.205

Adjusted R2 0.175 0.264 0.181

Residual Std. Error 73.026 (df = 464) 68.956 (df = 463) 72.737 (df = 463)

F Statistic 8.763\*\*\* (df = 13; 464) 13.225\*\*\* (df = 14; 463) 8.537\*\*\* (df = 14; 463)

===============================================================================================

**3. Select the best model from each set and examine whether it meets the assumptions of the regression model. Which of the five regression assumptions are met for the final models?**

|  |  |  |
| --- | --- | --- |
| **Assumption** | **M3: For deciding selling price** | **M12: For deciding days required to sell the house** |
| **Normality** | Observation: Residuals are comparatively normal for model  Result: Passed | Observation: Residuals are partially right skewed  Result: Failed |
| **Linearity** | Observation: Strong existence of outliers  Result: Failed | Observation: Dominant existence of high-end outliers  Result: Failed |
| **Homoskedasticity** | Observation: Heteroskedastic pattern is visible  Result: Failed | Observation: Heteroskedastic pattern is visible.  Result: Failed |
| **Multicollinearity** | GVIF Df GVIF^(1/(2\*Df))  age              1.225971  1        1.107236  sqft             4.750714  1        2.179613  bathsfull       52.796501  1        7.266120  I(bathsfull^2)  55.891492  1        7.476061  bathshalf       11.642673  1        3.412136  I(bathshalf^2)  11.009869  1        3.318112  garages        110.162842  1       10.495849  I(garages^2)   116.890291  1       10.811581  roof             1.758645  6        1.048169  splsale          1.193204  3        1.029878  beds             2.189388  1        1.479658  Observation: Many variables with square transformation have high vif which is natural. Otherwise, all other values are <10.  Result: Passed | GVIF Df GVIF^(1/(2\*Df))  listprice      25.662679  1        5.065834  I(listprice^2) 15.976183  1        3.997022  age             1.285644  1        1.133862  sqft            9.740239  1        3.120936  splsale         1.193777  3        1.029960  beds           76.043747  1        8.720307  I(beds^2)      73.937262  1        8.598678  bathsfull      49.825344  1        7.058707  I(bathsfull^2) 49.048033  1        7.003430  bathshalf      11.752102  1        3.428134  I(bathshalf^2) 11.001005  1        3.316776  garages         2.028653  1        1.424308  Observation: Many variables with square transformation have high vif which is natural. Otherwise, all other values are <10.  Result: Passed |
| **Independence** | Observation: No particular pattern in data. Any dependence between data points is not visible.  Result: Passed | Observation: No particular pattern in data. Any dependence between data points is not visible.  Result: Passed |

1. **Using your best models, select the top three predictors of adom and pricesold, and explain their marginal effects on the dependent variables. Remember that we are interested in economic significance, not statistical significance.**
   1. **Significant Variables for selling price variable –**

|  |  |
| --- | --- |
| **Variable** | **Impact** |
| Pool | Private pool increases value of house by $18229 |
| Area | Value of house increases by $136/sqft |
| \*RoofTile | Shake roof increase the value of house by $37178. |

\*Even though concrete roof and shake roof have more economical significance($120K more and $46K more resp.) they are not considered in table because they have very less observations in table.

* 1. **Significant variables for days required to sell house variable -**

|  |  |
| --- | --- |
| **Variable** | **Impact** |
| Pool | Houses with private pool take 11 days more than average to sell. |
| Bathshalf | Houses with powder room gets sold 12 days earlier than those without it |
| Splsale/shortsale | Disputed houses which are short sold by banks take 3 months more to close the sale than other houses |