**PRASAD GANESH HEGDE**

**(U76885007)**

**Final Sell Price**

|  |  |  |
| --- | --- | --- |
| **Predictor** | **Effect** | **Reason** |
| Beds | + | Sales price increases with increasing in number of bedrooms |
| Sqft | + | Price increases with increase in area of home, bigger houses tend to have high selling price |
| garages | + | As number of garages increases sale price increases up to certain number of garages, above certain limit number of garages do not affect the sales price |
| lotsqft | + | As lot area increase it increases the price |
| yrblt | - | Older the house is lesser is the selling price |
| Cdom | - | Longer the house sits on Market, decreases selling price. |

**New Features**

|  |  |  |
| --- | --- | --- |
| totalbath | + | combing the variable bathsfull + 0.5 \* bathhalf |
| rooftile | + | creating new binary variable for roof considering 1 for tile and 0 for Shingles roof, Tile roofs are more valued than shingle roof. |
| privatepool | + | separating private pool from Pool feature. house with private pool tends to have more selling price than community pool |
| pyear | +/- | as we need only year from pending to see the effect of trend, we are separating year from PendingDate variable |
| specialsale | - | Converting splsale to 1 if value is not none and 0 if it is none. Selling price decreases if house is listed in market by bank |

**Interaction Term**

|  |  |  |
| --- | --- | --- |
| Garage^2 | + | Till certain point number of garages have effect on selling price. To check the concave effect(saturation) creating this interaction |
| totalbath^2 | + | To check the saturation effect creating this interaction |
| sqft \* beds | + | As number of bedrooms increases area of the house increases and increases the price |

**Dropped Variables**

|  |
| --- |
| **listprice** – As it is highly correlated to sqft > 0.9. **lppersqft** – it’s function of list price and sqft |

**Adom**

|  |  |  |
| --- | --- | --- |
| Beds | + | number of bedrooms increases price, which in turn effect the number of potential buying customers so increases adom. |
| Garages | + | number of garages increases price, which in turn effect the number of potential buying customers so increases adom |
| Yrblt | - | newer is the house, less time to sell |
| specialsale | - | special houses have low listing price so tend to sell faster |
| Listprice | + | Higher the list price, effect the number of potential buying customer so increases adom |
| privatepool | - | People prefer Private pool, so private pool decreases the ADOM |
| totalbath | + | Number of total bath increases selling price, so increases Adom |
| Rooftile | + | Tile roof selling price will be higher than Shingles roof, so it increases adom |
| pyear | +/- | As market prices changes every year, depending on up and down in listing price on market adom increases or decreases |

**Interaction Terms**

|  |  |  |
| --- | --- | --- |
| Totalbath^2 | + | Up to certain number it has effect, after a limit it reaches saturation |
| Garage^2 | + | Up to certain number it has effect, after a limit it reaches saturation |

**m1 = lm(pricesold~beds+sqft+garages+lotsqft+yrblt+cdom\_cumuldaysmls+totalbath+**

**rooftile+privatepool+specialsale+pyear, data = df\_final\_t)**

**m2 = lm(pricesold~beds+sqft+garages+I(garages^2)+lotsqft+yrblt+cdom\_cumuldaysmls+totalbath+**

**I(totalbath^2)+rooftile+privatepool+specialsale+pyear, data = df\_final\_t)**

**m3 =lm(pricesold~beds+sqft++I(sqft\*beds)+garages+I(garages^2)+lotsqft+yrblt+cdom\_cumuldaysmls+totalbath+**

**I(totalbath^2)+rooftile+privatepool+specialsale+pyear, data = df\_final\_t)**

Text

Description automatically generated with low confidence

**m4 = lm(adom\_agentdaysonmarket~Beds+garages+yrblt+specialsale+listprice+**

**privatepool+totalbath+rooftile, data = df\_final)**

**m5 = lm(adom\_agentdaysonmarket~Beds+garages+yrblt+specialsale+listprice+**

**privatepool+totalbath+rooftile+I(totalbath^2), data = df\_final)**

**m6 = lm(adom\_agentdaysonmarket~Beds+garages+yrblt+specialsale+listprice+**

**privatepool+totalbath+rooftile+I(totalbath^2)+I(garages^2)+I(Beds^2), data = df\_final)**

Text

Description automatically generated

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?**

Checking the assumptions for m3 and m6

**Linearity**:

**m3: Does not follows linearity m6: does not follows linearity**

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

We can see pattern in residuals and residuals are not concentrated on 0 axis.

**Normality**:

**m3: does not follows normality m6: does not follows normality**

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

At tail and head data is not on qq line most of the points are not on qq line

**Equality of Variance**

**m3: does not follows Equality of variance m6: does not follows equality of variance**

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

Increasing pattern in Standardized residuals in both m3 and m6 model

**Independence**

m3: dose not follows Independent

durbinWatsonTest(m3)

#p-value is 0.016 so we reject null hypothesis, there is an auto correlation

m6: Does follows Independent

durbinWatsonTest(m6)

#p-value is 0.364 and greater than 0.05 so, we failed to reject null hypothesis there is no auto correlation

**Multicollinearity**

m3: does have multicollinearity problem

vif(m3)

beds sqft I(sqft \* beds) garages

14.898967 52.694439 100.518544 25.598703

I(garages^2) lotsqft yrblt cdom\_cumuldaysmls

30.134338 3.331715 1.289202 1.314778

totalbath I(totalbath^2) rooftile privatepool

54.883296 62.895179 1.535882 1.645300

specialsale pyear

1.031847 1.077120

For many of the features vif value is greater than 5.

m6: has multicollinearity problem, since vif values for many features are greater than 5.

Beds garages yrblt specialsale listprice

81.061925 25.313195 1.266851 1.067563 3.984133

privatepool totalbath rooftile I(totalbath^2) I(garages^2)

1.636532 37.679402 1.501589 41.486913 29.627539

I(Beds^2)

79.276611

**Selling price model do not meet any regression assumptions.**

**Adom model meets independent assumption.**

3. **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**

|  |  |  |
| --- | --- | --- |
| **m3** | **Beds** | **If one bedroom is added, it decreases selling price on an average by $79290 keeping all other factor constant.** |
| **m3** | **Privatepool** | **Private pool increases the selling price on an average by $21300 compared to community pool** |
| **m3** | **specialsale** | **Bank owned house decreases selling price on an average by $65650 compared to private owned house keeping all other factor constant.** |
| **m6** | **privatepool** | **House with private pool decreases adom by 9 days (approximately) on an average compared to house with community pool keeping all other factor constant.** |
| **m6** | **totalbath** | **For every addition of totalbath adom decreases by 7 days (approximately) on an average keeping all other factor constant.** |
| **m6** | **rooftile** | **House with tile roof decreases adom by 7 days(approximately) on an average compared to Shingle roof keeping all other factor constant.** |