Alexander Bukharin

5) We use the linear model function in R to fit the regression model. However, before we fit the regression, we format the data differently. Because location is a categorical variable, for each location we add a dummy variable that is equal to 1 if the location matches and 0 otherwise. However, we do this only for n - 1 locations, because you can encode which location it is with only n – 1 binary variables. Then we divide the data into 3 subsets: “foreclosure”, “short sale” and “regular” sales. We then fit the linear model for each subset. Here are the resulting coefficients. Because there are so many variables, we also use the step forward regression method to select which variables to keep for each type of sale. Below are the results for each variable type:

**Regular:**

Call:

lm(formula = regular$price ~ Bedrooms + Size + Price.SQ.Ft +

Oceano + Arroyo.Grande + Out.Of.Area.1 + Lompoc.1 + Santa.Maria.Orcutt.1,

data = regular)

Residuals:

Min 1Q Median 3Q Max

-507484 -104851 -25580 91767 1395237

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -914013.6 99879.9 -9.151 1.17e-14 \*\*\*

Bedrooms -56908.4 39059.7 -1.457 0.148460

Size 492.9 41.9 11.764 < 2e-16 \*\*\*

Price.SQ.Ft 2376.8 156.8 15.159 < 2e-16 \*\*\*

Oceano -1012312.3 271985.8 -3.722 0.000337 \*\*\*

Arroyo.Grande 168477.0 73071.9 2.306 0.023330 \*

Out.Of.Area.1 266641.4 142004.4 1.878 0.063523 .

Lompoc.1 385487.0 237897.5 1.620 0.108498

Santa.Maria.Orcutt.1 139378.0 73302.7 1.901 0.060312 .

---

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

Residual standard error: 233000 on 94 degrees of freedom

Multiple R-squared: 0.8902, Adjusted R-squared: 0.8809

F-statistic: 95.3 on 8 and 94 DF, p-value: < 2.2e-16

**Short Sale:**

Call:

lm(formula = regular$price ~ MLS + Bedrooms + Bathrooms + Size +

Price.SQ.Ft + Santa.Maria.Orcutt + Avila.Beach + Santa.Ynez +

Morro.Bay.1 + Bradley.1 + Cambria.1 + Oceano.1 + Cayucos +

Arroyo.Grande.1 + Nipomo.1, data = regular)

Residuals:

Min 1Q Median 3Q Max

-737496 -20960 5230 26667 538493

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -5.240e+04 1.246e+05 -0.420 0.674329

MLS -1.361e+00 8.132e-01 -1.673 0.094910 .

Bedrooms -1.908e+04 5.627e+03 -3.391 0.000752 \*\*\*

Bathrooms 1.564e+04 6.286e+03 2.487 0.013199 \*

Size 2.050e+02 7.012e+00 29.238 < 2e-16 \*\*\*

Price.SQ.Ft 1.339e+03 4.963e+01 26.977 < 2e-16 \*\*\*

Santa.Maria.Orcutt -1.343e+04 8.112e+03 -1.656 0.098431 .

Avila.Beach 6.651e+05 7.873e+04 8.448 3.26e-16 \*\*\*

Santa.Ynez 2.116e+05 3.971e+04 5.328 1.50e-07 \*\*\*

Morro.Bay.1 5.365e+04 2.812e+04 1.908 0.056974 .

Bradley.1 -9.161e+04 5.461e+04 -1.677 0.094081 .

Cambria.1 -1.226e+05 3.588e+04 -3.417 0.000685 \*\*\*

Oceano.1 6.477e+04 2.938e+04 2.205 0.027918 \*

Cayucos 4.403e+05 7.801e+04 5.644 2.79e-08 \*\*\*

Arroyo.Grande.1 3.725e+04 1.707e+04 2.182 0.029580 \*

Nipomo.1 3.008e+04 1.562e+04 1.925 0.054746 .

---

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

Residual standard error: 76030 on 500 degrees of freedom

Multiple R-squared: 0.8997, Adjusted R-squared: 0.8967

F-statistic: 299 on 15 and 500 DF, p-value: < 2.2e-16

**Foreclosure:**

Call:

lm(formula = regular$price ~ Bedrooms + Bathrooms + Size + Price.SQ.Ft +

Avila.Beach + Paso.Robles + Pismo.Beach + San.Luis.Obispo +

Santa.Margarita + King.City, data = regular)

Residuals:

Min 1Q Median 3Q Max

-280038 -17002 3121 18865 280038

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.349e+05 2.004e+04 -16.712 < 2e-16 \*\*\*

Bedrooms -1.011e+04 6.152e+03 -1.642 0.1026

Bathrooms 1.603e+04 8.654e+03 1.852 0.0659 .

Size 2.035e+02 8.097e+00 25.133 < 2e-16 \*\*\*

Price.SQ.Ft 1.701e+03 5.958e+01 28.558 < 2e-16 \*\*\*

Avila.Beach -2.423e+05 4.844e+04 -5.002 1.56e-06 \*\*\*

Paso.Robles -2.078e+04 1.210e+04 -1.717 0.0881 .

Pismo.Beach 2.476e+05 4.178e+04 5.927 2.02e-08 \*\*\*

San.Luis.Obispo -6.549e+04 2.609e+04 -2.510 0.0131 \*

Santa.Margarita 9.882e+04 5.623e+04 1.757 0.0809 .

King.City -6.752e+04 3.305e+04 -2.043 0.0428 \*

---

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

Residual standard error: 54720 on 151 degrees of freedom

Multiple R-squared: 0.9615, Adjusted R-squared: 0.959

F-statistic: 377.2 on 10 and 151 DF, p-value: < 2.2e-16