Predicting sale price given other variabes:

Variables:

* Numerical:
  + Sqfeet
  + Year
  + Lot
* Cat:
  + Beds
  + Baths
  + Air conditioning
  + Garage
  + Pool
  + Quality
  + Style
  + Highway

Binary Variables:

* Air
* Pool
* highway

Categorical Variables:

* Quality (3)
* Garage(7)
  + Combine 3 4 5 7
* Baths(7)
  + Combine 5 6 7
* Beds(7)
  + Just combine 6 and 7
* Style(10) 🡪 4 groups
  + 1+2
  + 3+4
  + 5+6
  + 7+

Regression:

* Full Model Results:

|  |
| --- |
| Call:  lm(formula = sqfeet ~ ., data = reg\_data)  Residuals:  Min 1Q Median 3Q Max  -0.87649 -0.19301 -0.01171 0.15458 1.68374  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 4.707662 2.347703 2.005 0.045480 \*  year -0.001330 0.001193 -1.115 0.265448  lot 0.002806 0.001367 2.053 0.040618 \*  highway1 0.025592 0.104204 0.246 0.806096  air1 0.044434 0.047736 0.931 0.352397  pool1 0.040094 0.060726 0.660 0.509407  quality2 -0.563536 0.057850 -9.741 < 2e-16 \*\*\*  quality3 -0.762939 0.078153 -9.762 < 2e-16 \*\*\*  style\_g42 0.063062 0.045682 1.380 0.168064  style\_g43 0.430424 0.064130 6.712 5.25e-11 \*\*\*  style\_g44+ 0.509038 0.042915 11.862 < 2e-16 \*\*\*  beds\_g62 0.048817 0.121740 0.401 0.688597  beds\_g63 0.123433 0.120266 1.026 0.305230  beds\_g64 0.215517 0.122546 1.759 0.079250 .  beds\_g65 0.306599 0.131525 2.331 0.020146 \*  beds\_g66+ 0.604878 0.154700 3.910 0.000105 \*\*\*  baths\_g52 0.075050 0.054100 1.387 0.165989  baths\_g53 0.240389 0.068633 3.503 0.000502 \*\*\*  baths\_g54 0.476265 0.083916 5.675 2.35e-08 \*\*\*  baths\_g55+ 0.843487 0.112952 7.468 3.68e-13 \*\*\*  garage\_g41 -0.035134 0.139384 -0.252 0.801094  garage\_g42 0.018581 0.135703 0.137 0.891145  garage\_g43+ 0.188643 0.142935 1.320 0.187513  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.3352 on 498 degrees of freedom  Multiple R-squared: 0.7875, Adjusted R-squared: 0.7781  F-statistic: 83.9 on 22 and 498 DF, p-value: < 2.2e-16 |
|  |
| |  | | --- | | > | |

Took away highway:

Call:

lm(formula = sqfeet ~ year + lot + air + pool + quality + style\_g4 +

beds\_g6 + baths\_g5 + garage\_g4, data = reg\_data)

Residuals:

Min 1Q Median 3Q Max

-0.87529 -0.19385 -0.01237 0.15510 1.68315

Coefficients:

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

(Intercept) 4.660724 2.337707 1.994 0.046727 \*

year -0.001306 0.001188 -1.100 0.272039

lot 0.002835 0.001361 2.083 0.037768 \*

air1 0.043712 0.047601 0.918 0.358904

pool1 0.039762 0.060654 0.656 0.512413

quality2 -0.563754 0.057789 -9.755 < 2e-16 \*\*\*

quality3 -0.763140 0.078075 -9.775 < 2e-16 \*\*\*

style\_g42 0.062600 0.045600 1.373 0.170430

style\_g43 0.429669 0.063996 6.714 5.17e-11 \*\*\*

style\_g44+ 0.508276 0.042762 11.886 < 2e-16 \*\*\*

beds\_g62 0.049494 0.121594 0.407 0.684151

beds\_g63 0.123823 0.120142 1.031 0.303209

beds\_g64 0.215889 0.122422 1.763 0.078431 .

beds\_g65 0.307345 0.131366 2.340 0.019697 \*

beds\_g66+ 0.605531 0.154531 3.919 0.000102 \*\*\*

baths\_g52 0.075394 0.054031 1.395 0.163520

baths\_g53 0.240369 0.068568 3.506 0.000497 \*\*\*

baths\_g54 0.475774 0.083813 5.677 2.33e-08 \*\*\*

baths\_g55+ 0.842579 0.112785 7.471 3.60e-13 \*\*\*

garage\_g41 -0.034995 0.139252 -0.251 0.801683

garage\_g42 0.019513 0.135522 0.144 0.885574

garage\_g43+ 0.189114 0.142787 1.324 0.185960

---

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

Residual standard error: 0.3349 on 499 degrees of freedom

Multiple R-squared: 0.7875, Adjusted R-squared: 0.7785

F-statistic: 88.05 on 21 and 499 DF, p-value: < 2.2e-16

-garage

Call:

lm(formula = sqfeet ~ year + lot + air + pool + quality + style\_g4 +

beds\_g6 + baths\_g5, data = reg\_data)

Residuals:

Min 1Q Median 3Q Max

-0.93861 -0.20059 -0.03429 0.16397 1.67107

Coefficients:

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

**(Intercept) 3.1224172 2.3276430 1.341 0.180381**

year -0.0004819 0.0011795 -0.409 0.683048

lot 0.0035402 0.0013634 2.597 0.009693 \*\*

air1 0.0537467 0.0475174 1.131 0.258555

pool1 0.0410650 0.0613101 0.670 0.503299

quality2 -0.6433520 0.0539017 -11.936 < 2e-16 \*\*\*

quality3 -0.8525096 0.0751615 -11.342 < 2e-16 \*\*\*

style\_g42 0.0660048 0.0457207 1.444 0.149460

style\_g43 0.4373119 0.0645708 6.773 3.55e-11 \*\*\*

style\_g44+ 0.5211006 0.0430742 12.098 < 2e-16 \*\*\*

beds\_g62 0.0526609 0.1211644 0.435 0.664022

beds\_g63 0.1251978 0.1182877 1.058 0.290374

beds\_g64 0.2168073 0.1206749 1.797 0.072996 .

beds\_g65 0.3117493 0.1300237 2.398 0.016866 \*

beds\_g66+ 0.6118383 0.1538930 3.976 8.05e-05 \*\*\*

baths\_g52 0.0746811 0.0537147 1.390 0.165045

baths\_g53 0.2502860 0.0685100 3.653 0.000286 \*\*\*

baths\_g54 0.5027778 0.0840518 5.982 4.20e-09 \*\*\*

baths\_g55+ 0.8446504 0.1135379 7.439 4.42e-13 \*\*\*

---

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

Residual standard error: 0.3388 on 502 degrees of freedom

Multiple R-squared: 0.7813, Adjusted R-squared: 0.7734

F-statistic: 99.63 on 18 and 502 DF, p-value: < 2.2e-16

-year:

Call:

lm(formula = sqfeet ~ lot + air + pool + quality + style\_g4 +

beds\_g6 + baths\_g5, data = reg\_data)

Residuals:

Min 1Q Median 3Q Max

-0.93580 -0.20109 -0.03447 0.16783 1.68050

Coefficients:

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

(Intercept) 2.173246 0.142074 15.297 < 2e-16 \*\*\*

lot 0.003649 0.001336 2.731 0.006540 \*\*

air1 0.050222 0.046689 1.076 0.282590

pool1 0.043019 0.061073 0.704 0.481515

quality2 -0.638069 0.052284 -12.204 < 2e-16 \*\*\*

quality3 -0.841984 0.070549 -11.935 < 2e-16 \*\*\*

style\_g42 0.065616 0.045673 1.437 0.151440

style\_g43 0.443000 0.062999 7.032 6.68e-12 \*\*\*

style\_g44+ 0.521155 0.043038 12.109 < 2e-16 \*\*\*

beds\_g62 0.049938 0.120881 0.413 0.679695

beds\_g63 0.122782 0.118042 1.040 0.298767

beds\_g64 0.213846 0.120357 1.777 0.076211 .

beds\_g65 0.310109 0.129854 2.388 0.017302 \*

beds\_g66+ 0.610169 0.153711 3.970 8.25e-05 \*\*\*

baths\_g52 0.073375 0.053575 1.370 0.171430

baths\_g53 0.247457 0.068103 3.634 0.000308 \*\*\*

baths\_g54 0.497023 0.082795 6.003 3.71e-09 \*\*\*

baths\_g55+ 0.837045 0.111909 7.480 3.34e-13 \*\*\*

---

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

Residual standard error: 0.3385 on 503 degrees of freedom

Multiple R-squared: 0.7812, Adjusted R-squared: 0.7738

F-statistic: 105.7 on 17 and 503 DF, p-value: < 2.2e-16

#-pool

Call:

lm(formula = sqfeet ~ lot + air + quality + style\_g4 + beds\_g6 +

baths\_g5, data = reg\_data)

Residuals:

Min 1Q Median 3Q Max

-0.94314 -0.19544 -0.03649 0.16629 1.67713

Coefficients:

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

(Intercept) 2.172997 0.142003 15.303 < 2e-16 \*\*\*

lot 0.003589 0.001333 2.693 0.007316 \*\*

air1 0.051710 0.046618 1.109 0.267861

quality2 -0.637704 0.052256 -12.204 < 2e-16 \*\*\*

quality3 -0.841412 0.070509 -11.933 < 2e-16 \*\*\*

style\_g42 0.064145 0.045602 1.407 0.160156

style\_g43 0.441828 0.062946 7.019 7.24e-12 \*\*\*

style\_g44+ 0.520014 0.042986 12.097 < 2e-16 \*\*\*

beds\_g62 0.050520 0.120817 0.418 0.676016

beds\_g63 0.124931 0.117943 1.059 0.289997

beds\_g64 0.214447 0.120294 1.783 0.075238 .

beds\_g65 0.309842 0.129789 2.387 0.017341 \*

beds\_g66+ 0.622924 0.152565 4.083 5.17e-05 \*\*\*

baths\_g52 0.073914 0.053543 1.380 0.168054

baths\_g53 0.250324 0.067947 3.684 0.000254 \*\*\*

baths\_g54 0.501669 0.082490 6.082 2.35e-09 \*\*\*

baths\_g55+ 0.844991 0.111283 7.593 1.53e-13 \*\*\*

---

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

Residual standard error: 0.3383 on 504 degrees of freedom

Multiple R-squared: 0.781, Adjusted R-squared: 0.774

F-statistic: 112.3 on 16 and 504 DF, p-value: < 2.2e-16

-air

Call:

lm(formula = sqfeet ~ lot + quality + style\_g4 + beds\_g6 + baths\_g5,

data = reg\_data)

Residuals:

Min 1Q Median 3Q Max

-0.9382 -0.2041 -0.0345 0.1730 1.6816

Coefficients:

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

(Intercept) 2.218304 0.136032 16.307 < 2e-16 \*\*\*

lot 0.003314 0.001310 2.530 0.011700 \*

quality2 -0.641995 0.052124 -12.317 < 2e-16 \*\*\*

quality3 -0.859727 0.068565 -12.539 < 2e-16 \*\*\*

style\_g42 0.063980 0.045613 1.403 0.161326

style\_g43 0.432265 0.062367 6.931 1.28e-11 \*\*\*

style\_g44+ 0.517776 0.042949 12.056 < 2e-16 \*\*\*

beds\_g62 0.051715 0.120840 0.428 0.668861

beds\_g63 0.131106 0.117839 1.113 0.266416

beds\_g64 0.220207 0.120209 1.832 0.067561 .

beds\_g65 0.320974 0.129429 2.480 0.013466 \*

beds\_g66+ 0.631135 0.152420 4.141 4.06e-05 \*\*\*

baths\_g52 0.084033 0.052772 1.592 0.111927

baths\_g53 0.258710 0.067540 3.830 0.000144 \*\*\*

baths\_g54 0.511847 0.081997 6.242 9.15e-10 \*\*\*

baths\_g55+ 0.849022 0.111249 7.632 1.16e-13 \*\*\*

---

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

Residual standard error: 0.3384 on 505 degrees of freedom

Multiple R-squared: 0.7805, Adjusted R-squared: 0.7739

F-statistic: 119.7 on 15 and 505 DF, p-value: < 2.2e-16