Step 1: Fit full model.

Call:

lm(formula = Price ~ ., data = sampled.data)

Residuals:

Min 1Q Median 3Q Max

-170551 -38278 -2521 31102 312232

Coefficients:

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

(Intercept) -1.714e+06 5.649e+05 -3.035 0.002626 \*\*

SqFt 1.224e+02 9.778e+00 12.516 < 2e-16 \*\*\*

Bed -1.067e+04 4.314e+03 -2.474 0.013943 \*

Bath 3.358e+03 5.616e+03 0.598 0.550354

AC -2.054e+03 1.106e+04 -0.186 0.852844

GSize 1.709e+04 7.201e+03 2.373 0.018302 \*

Pool 5.722e+03 1.426e+04 0.401 0.688451

Year 9.256e+02 2.855e+02 3.242 0.001327 \*\*

Quality -5.242e+04 9.370e+03 -5.594 5.15e-08 \*\*\*

Style -7.232e+03 1.922e+03 -3.763 0.000204 \*\*\*

LotSize 1.158e+00 3.259e-01 3.554 0.000444 \*\*\*

Hwy -1.980e+04 2.831e+04 -0.699 0.484917

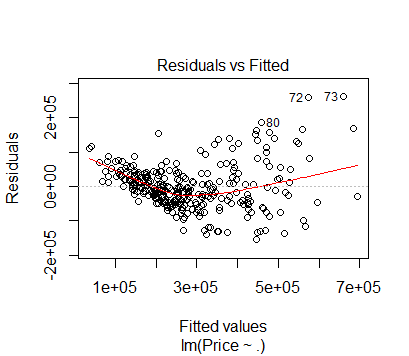
---

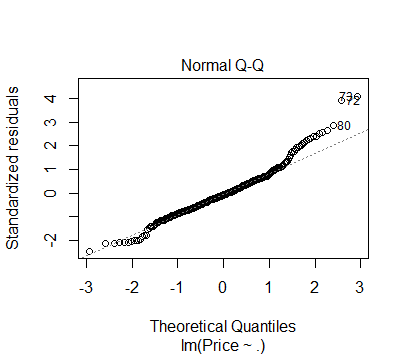
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 61880 on 288 degrees of freedom

Multiple R-squared: 0.8085, Adjusted R-squared: 0.8012

F-statistic: 110.5 on 11 and 288 DF, p-value: < 2.2e-16





**OBSERVATIONS**

Hypothesis testing for all variables:

At alpha 0.05 level of significance,

p-value > alpha for Bath, AC, Pool, Hwy i.e. we fail to reject null hypothesis, hence beta co-efficient of these predictors are equal to zero. Hence these seem to be like useless predictors or other predictors are making them appear useless. This requires check for multi-co-linearity.

We have sufficient evidence at alpha 0.5 level of significance that at least one predictor is useful, since the overall p-value (2.2e-16) is less than alpha.

Negative slope for Bed, AC, quality, style, Hwy.

There seems to a U shape in the residual vs fitted plot.

This means there is problems with linearity in data.

From Q-Q plot, the central data points look linear but towards the tail up & down, data points are far from the linear line.

*PROBLEM: DATA LINEARITY*

Step 2: Find the co relation between all the variables

Price SqFt Bed Bath AC GSize Pool Year

Price 1.00000000 0.8070805 0.37208172 0.66578106 0.308661889 0.54282450 0.16153797 0.56274065

SqFt 0.80708045 1.0000000 0.51222206 0.75879598 0.264659922 0.46844325 0.18232928 0.40659904

Bed 0.37208172 0.5122221 1.00000000 0.62447206 0.255244975 0.25934688 0.17520632 0.23098086

Bath 0.66578106 0.7587960 0.62447206 1.00000000 0.321256203 0.43280748 0.25986782 0.47592697

AC 0.30866189 0.2646599 0.25524498 0.32125620 1.000000000 0.30782739 0.09790468 0.45273153

GSize 0.54282450 0.4684432 0.25934688 0.43280748 0.307827389 1.00000000 0.04714343 0.49624460

Pool 0.16153797 0.1823293 0.17520632 0.25986782 0.097904675 0.04714343 1.00000000 0.05308074

Year 0.56274065 0.4065990 0.23098086 0.47592697 0.452731528 0.49624460 0.05308074 1.00000000

Quality -0.74095005 -0.6656879 -0.33082019 -0.64460578 -0.433043077 -0.54509041 -0.14294184 -0.65030423

Style 0.33448344 0.6190403 0.34430136 0.48160859 0.121114038 0.25091064 0.12301800 0.21112018

LotSize 0.17238892 0.1301643 0.16052955 0.13958915 -0.096104990 0.15791140 -0.01573362 -0.13081965

Hwy -0.05623466 -0.0507693 -0.03611378 -0.04077068 0.007332706 0.01873091 -0.03919309 0.02647923

Quality Style LotSize Hwy

Price -0.740950053 0.33448344 0.172388922 -0.056234656

SqFt -0.665687917 0.61904025 0.130164330 -0.050769304

Bed -0.330820189 0.34430136 0.160529555 -0.036113777

Bath -0.644605781 0.48160859 0.139589148 -0.040770677

AC -0.433043077 0.12111404 -0.096104990 0.007332706

GSize -0.545090410 0.25091064 0.157911397 0.018730915

Pool -0.142941842 0.12301800 -0.015733623 -0.039193090

Year -0.650304230 0.21112018 -0.130819650 0.026479227

Quality 1.000000000 -0.31454177 0.008635689 -0.006576651

Style -0.314541772 1.00000000 0.036098082 -0.117463777

LotSize 0.008635689 0.03609808 1.000000000 0.067887415

Hwy -0.006576651 -0.11746378 0.067887415 1.000000000

Call:

lm(formula = Price ~ SqFt, data = sampled.data)

Residuals:

Min 1Q Median 3Q Max

-207171 -45483 -8748 25353 382107

Coefficients:

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

(Intercept) -81914.74 16020.24 -5.113 5.68e-07 \*\*\*

SqFt 160.70 6.81 23.597 < 2e-16 \*\*\*

---

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

Residual standard error: 84720 on 298 degrees of freedom

Multiple R-squared: 0.6514, Adjusted R-squared: 0.6502

F-statistic: 556.8 on 1 and 298 DF, p-value: < 2.2e-16