Multiple Linear Regression Development

James Davin, Rui Gong, Jiemin Yang

2024-03-29

Multiple linear regression

From our previous work with correlation matrices, we can see that the potentially most predictive elements for price included engine power, mileage, age, Feature 3, and Feature 8. However, other features had positive effects and we did not investigate the effect of paint and model.

We convert the model, color, paint, and month-sold and month-registered information all to categorical data for R to interpret as separate classes.

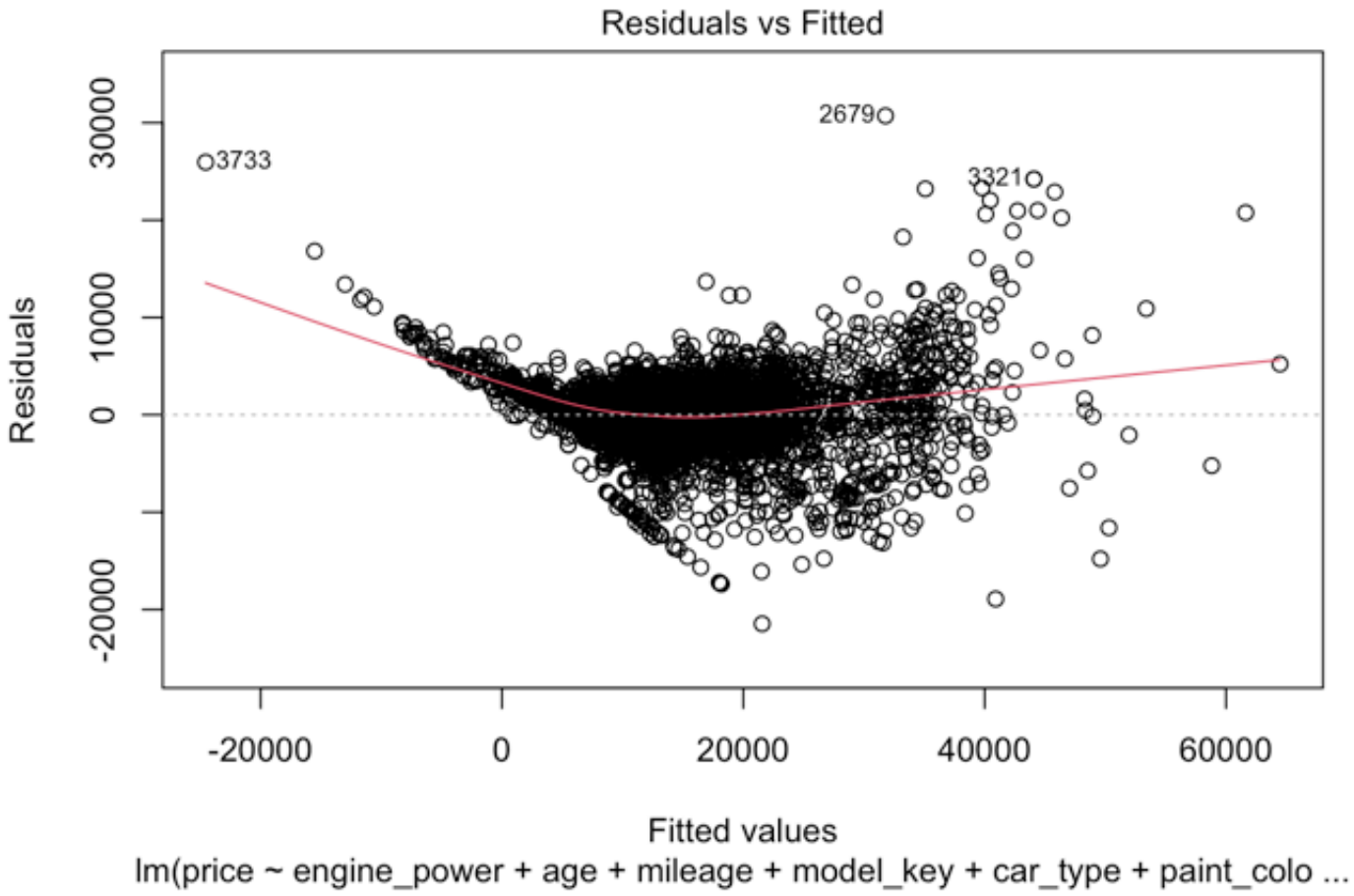
Initially, we construct a multiple linear regression model incorporating all potentially relevant features, excluding those with apparent multicollinearity. We also omit the month of registration from our analysis because it duplicates information captured by the age feature and appears to have an illogical impact on the subsequent selling price. Fitting the intial model, we will then analyze the model's summary to determine the most significant features.

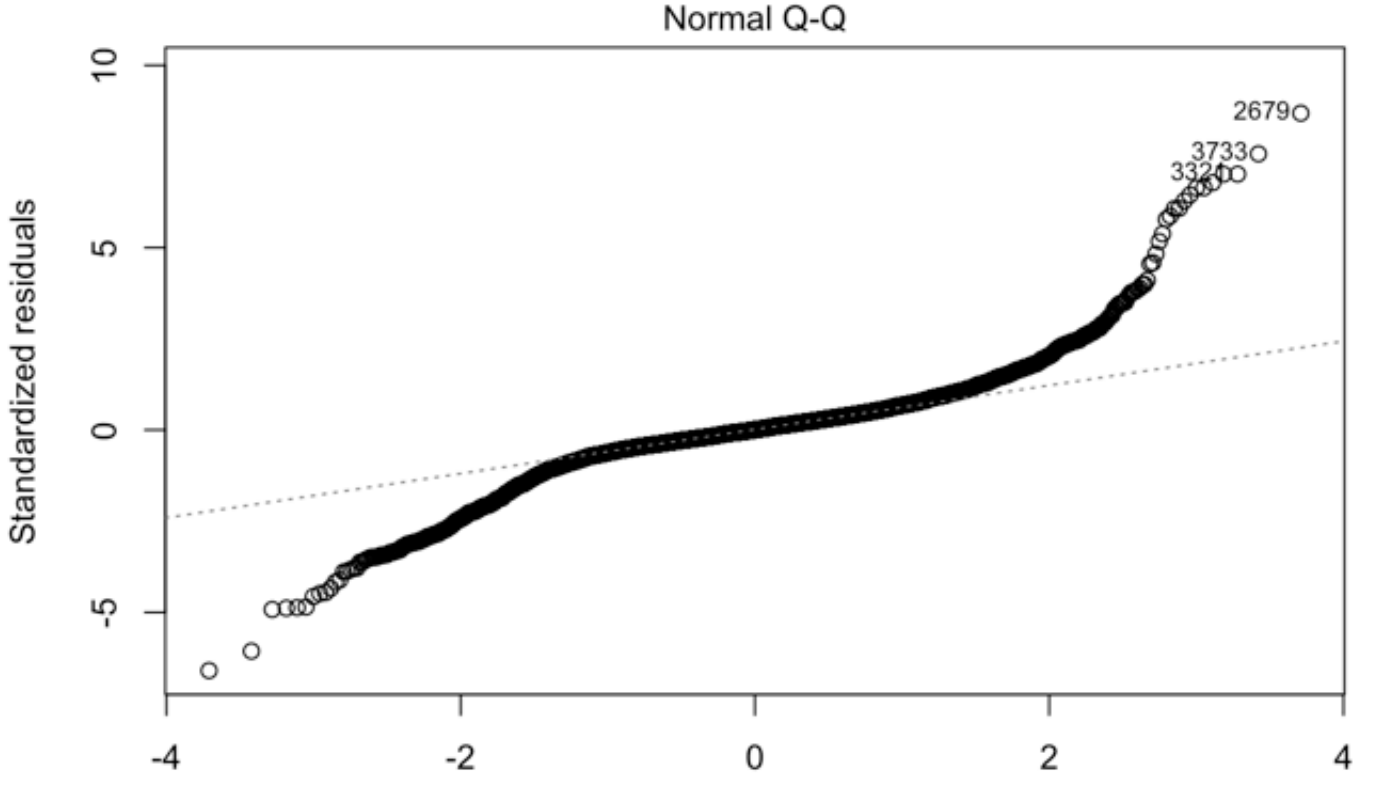
```
##
## Call:
## lm(formula = price ~ engine_power + age + mileage + model_key +
##     car_type + paint_color + month_sold + fuel + feature_1 +
##     feature_2 + feature_3 + feature_4 + feature_5 + feature_6 +
##     feature_7 + feature_8, data = bmw_data_cleaned)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21468.0  -1386.5     14.7   1480.4  30714.9
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.578e+04  1.384e+03  11.400  < 2e-16 ***
## engine_power                5.198e+01  3.965e+00  13.108  < 2e-16 ***
## age                        -1.059e+03  2.923e+01 -36.232  < 2e-16 ***
## mileage                    -3.095e-02  1.079e-03 -28.699  < 2e-16 ***
## model_key116                1.272e+02  9.493e+02   0.134 0.893386
## model_key118                4.101e+02  9.900e+02   0.414 0.678742
## model_key120                3.733e+02  1.144e+03   0.326 0.744243
## model_key123               -2.712e+03  2.716e+03  -0.999 0.318024
## model_key125               -2.376e+03  1.885e+03  -1.261 0.207452
## model_key135                1.447e+03  2.122e+03   0.682 0.495550
## model_key214 Gran Tourer    8.642e+03  3.970e+03   2.177 0.029531 *
## model_key216               -3.228e+03  3.934e+03  -0.821 0.411836
```

```
## model_key216 Active Tourer  6.130e+03  2.641e+03   2.322 0.020302 *
## model_key216 Gran Tourer    7.592e+03  2.433e+03   3.121 0.001815 **
## model_key218                2.920e+03  1.498e+03   1.949 0.051332 .
## model_key218 Active Tourer  6.717e+03  1.912e+03   3.512 0.000448 ***
## model_key218 Gran Tourer    5.089e+03  2.043e+03   2.490 0.012791 *
## model_key220                2.782e+03  2.400e+03   1.159 0.246476
## model_key220 Active Tourer  2.586e+03  3.942e+03   0.656 0.511883
## model_key225               -7.567e+02  3.759e+03  -0.201 0.840476
## model_key225 Active Tourer  5.908e+03  3.998e+03   1.478 0.139566
## model_key316                2.027e+03  1.025e+03   1.977 0.048113 *
## model_key318                2.508e+03  1.019e+03   2.460 0.013919 *
## model_key318 Gran Turismo   3.563e+03  1.018e+03   3.501 0.000468 ***
## model_key320                2.046e+03  1.029e+03   1.988 0.046887 *
## model_key320 Gran Turismo   3.327e+03  1.064e+03   3.127 0.001779 **
## model_key325                8.292e+02  1.338e+03   0.620 0.535482
## model_key325 Gran Turismo   6.415e+03  2.047e+03   3.134 0.001737 **
## model_key328                2.018e+03  1.935e+03   1.043 0.296974
## model_key330                1.500e+03  1.265e+03   1.185 0.235931
## model_key330 Gran Turismo   4.244e+03  2.323e+03   1.827 0.067760 .
## model_key335                3.702e+03  1.780e+03   2.080 0.037622 *
## model_key335 Gran Turismo   3.781e+03  2.352e+03   1.608 0.107896
## model_key418 Gran Coupé     6.515e+03  1.584e+03   4.112 3.99e-05 ***
## model_key420                7.299e+03  1.229e+03   5.940 3.05e-09 ***
## model_key420 Gran Coupé     7.065e+03  1.154e+03   6.120 1.01e-09 ***
## model_key425                1.169e+04  2.772e+03   4.218 2.51e-05 ***
## model_key430                8.185e+03  2.786e+03   2.937 0.003327 **
## model_key430 Gran Coupé     6.331e+03  2.092e+03   3.026 0.002493 **
## model_key435                7.037e+03  2.188e+03   3.216 0.001309 **
## model_key435 Gran Coupé     1.051e+04  1.834e+03   5.728 1.08e-08 ***
## model_key518                5.205e+03  1.106e+03   4.707 2.58e-06 ***
## model_key520                4.951e+03  1.041e+03   4.754 2.06e-06 ***
## model_key520 Gran Turismo   6.895e+03  1.232e+03   5.597 2.30e-08 ***
## model_key523                7.103e+03  2.095e+03   3.390 0.000704 ***
## model_key525                4.586e+03  1.095e+03   4.187 2.87e-05 ***
## model_key528                4.803e+03  1.827e+03   2.629 0.008581 **
## model_key530                5.115e+03  1.143e+03   4.474 7.87e-06 ***
## model_key530 Gran Turismo   5.654e+03  1.362e+03   4.152 3.35e-05 ***
## model_key535                5.405e+03  1.293e+03   4.180 2.97e-05 ***
## model_key535 Gran Turismo   3.775e+03  2.763e+03   1.366 0.171967
## model_key630               -2.968e+02  3.771e+03  -0.079 0.937260
## model_key635                5.532e+03  3.783e+03   1.462 0.143740
## model_key640                1.487e+04  1.805e+03   8.241  < 2e-16 ***
## model_key640 Gran Coupé     1.585e+04  1.423e+03  11.141  < 2e-16 ***
## model_key650               -1.157e+04  2.902e+03  -3.987 6.79e-05 ***
## model_key730                9.379e+03  1.267e+03   7.401 1.59e-13 ***
## model_key735                1.082e+03  3.764e+03   0.287 0.773752
## model_key740                1.706e+04  1.455e+03  11.724  < 2e-16 ***
## model_key750                1.115e+04  2.838e+03   3.930 8.61e-05 ***
## model_keyActiveHybrid 5    -3.392e+03  4.276e+03  -0.793 0.427635
## model_keyi3                 5.884e+03  3.083e+03   1.908 0.056399 .
```

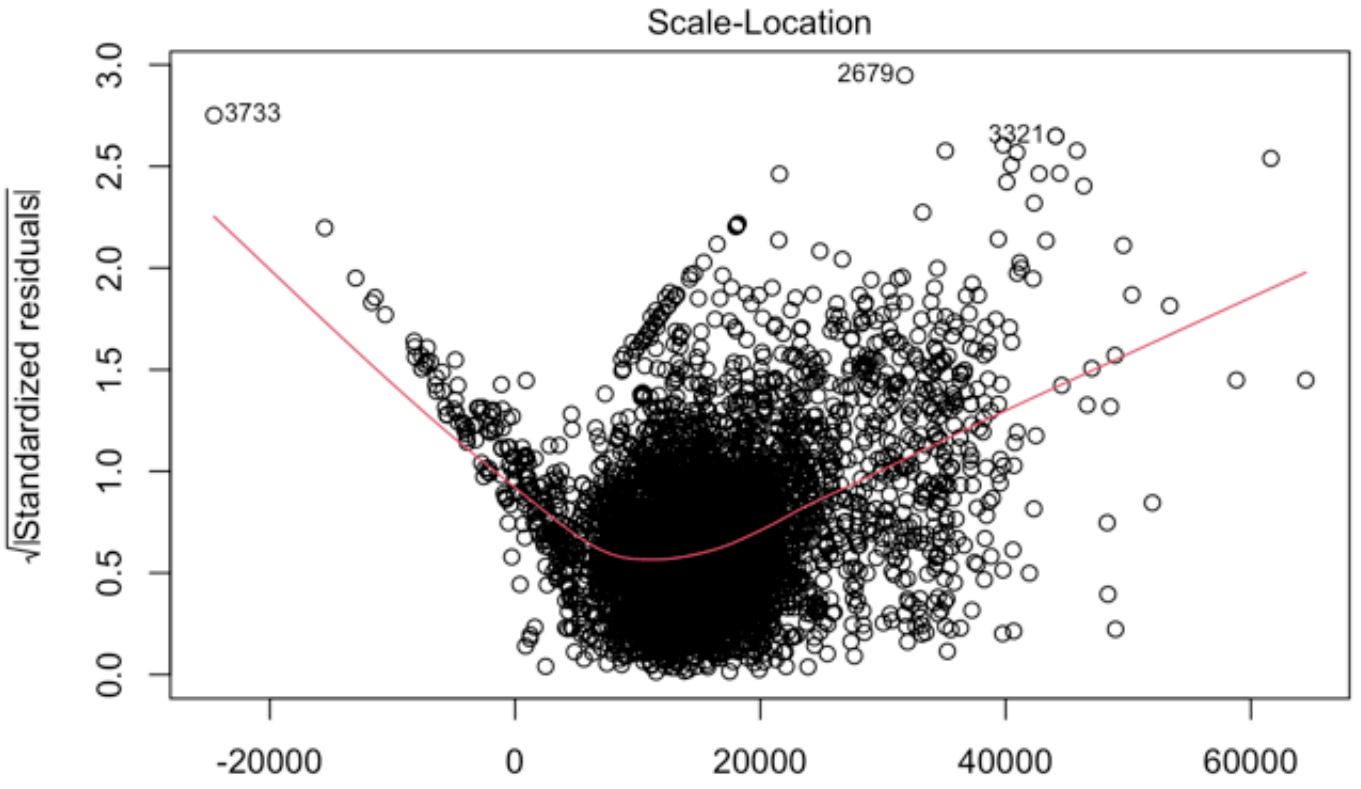
```
## model_keyM135               5.378e+03  3.750e+03   1.434 0.151526
## model_keyM235               5.152e+03  2.427e+03   2.123 0.033810 *
## model_keyM3                 1.675e+04  1.930e+03   8.679  < 2e-16 ***
## model_keyM4                 3.075e+04  2.904e+03  10.588  < 2e-16 ***
## model_keyM5                 1.331e+04  3.940e+03   3.378 0.000735 ***
## model_keyM550               1.123e+04  1.651e+03   6.801 1.17e-11 ***
## model_keyX1                 1.823e+03  1.907e+03   0.956 0.339187
## model_keyX3                 4.882e+03  1.910e+03   2.556 0.010622 *
## model_keyX4                 1.306e+04  1.997e+03   6.540 6.80e-11 ***
## model_keyX5                 1.379e+04  1.977e+03   6.976 3.46e-12 ***
## model_keyX5 M               1.747e+04  2.154e+03   8.111 6.35e-16 ***
## model_keyX5 M50             2.418e+04  2.931e+03   8.251  < 2e-16 ***
## model_keyX6                 1.594e+04  2.063e+03   7.727 1.33e-14 ***
## model_keyX6 M               2.600e+04  2.398e+03  10.844  < 2e-16 ***
## model_keyZ4                 3.431e+02  1.904e+03   0.180 0.857054
## car_typecoupe              -2.968e+03  7.277e+02  -4.078 4.62e-05 ***
## car_typeestate             -4.722e+03  6.910e+02  -6.833 9.37e-12 ***
## car_typehatchback          -3.360e+03  7.132e+02  -4.712 2.52e-06 ***
## car_typesedan              -2.708e+03  6.879e+02  -3.937 8.36e-05 ***
## car_typesubcompact         -2.330e+03  7.927e+02  -2.939 0.003304 **
## car_typesuv                -3.721e+03  1.720e+03  -2.164 0.030536 *
## car_typevan                -8.740e+03  1.439e+03  -6.075 1.34e-09 ***
## paint_colorblack            1.633e+02  5.806e+02   0.281 0.778492
## paint_colorblue            -2.751e+02  5.891e+02  -0.467 0.640506
## paint_colorbrown            1.689e+02  6.060e+02   0.279 0.780406
## paint_colorgreen           -9.804e+01  1.033e+03  -0.095 0.924400
## paint_colorgrey             4.775e+01  5.833e+02   0.082 0.934748
## paint_colororange          -7.311e+02  1.636e+03  -0.447 0.654960
## paint_colorred              5.131e+02  7.654e+02   0.670 0.502657
## paint_colorsilver          -8.310e+01  6.069e+02  -0.137 0.891085
## paint_colorwhite            5.527e+02  5.958e+02   0.928 0.353621
## month_sold2                -3.073e-01  2.965e+02  -0.001 0.999173
## month_sold3                -1.546e+02  2.824e+02  -0.547 0.584222
## month_sold4                 1.162e+02  2.852e+02   0.408 0.683654
## month_sold5                 2.490e+02  2.801e+02   0.889 0.374019
## month_sold6                 1.597e+02  2.903e+02   0.550 0.582185
## month_sold7                 4.882e+02  2.955e+02   1.652 0.098523 .
## month_sold8                 6.519e+02  2.974e+02   2.192 0.028416 *
## month_sold9                 5.290e+02  3.498e+02   1.512 0.130512
## fuelelectro                 1.099e+03  3.598e+03   0.305 0.760076
## fuelhybrid_petrol           4.559e+03  2.080e+03   2.192 0.028453 *
## fuelpetrol                 -4.210e+02  3.240e+02  -1.300 0.193802
## feature_1TRUE               5.780e+02  1.252e+02   4.618 3.98e-06 ***
## feature_2TRUE              -2.877e-01  1.572e+02  -0.002 0.998540
## feature_3TRUE               6.930e+02  1.409e+02   4.918 9.05e-07 ***
## feature_4TRUE               1.435e+03  1.752e+02   8.190 3.34e-16 ***
## feature_5TRUE              -1.498e+02  1.252e+02  -1.197 0.231547
## feature_6TRUE               1.320e+03  1.321e+02   9.991  < 2e-16 ***
## feature_7TRUE               1.015e+03  2.446e+02   4.149 3.39e-05 ***
## feature_8TRUE               1.311e+03  1.294e+02  10.133  < 2e-16 ***
```

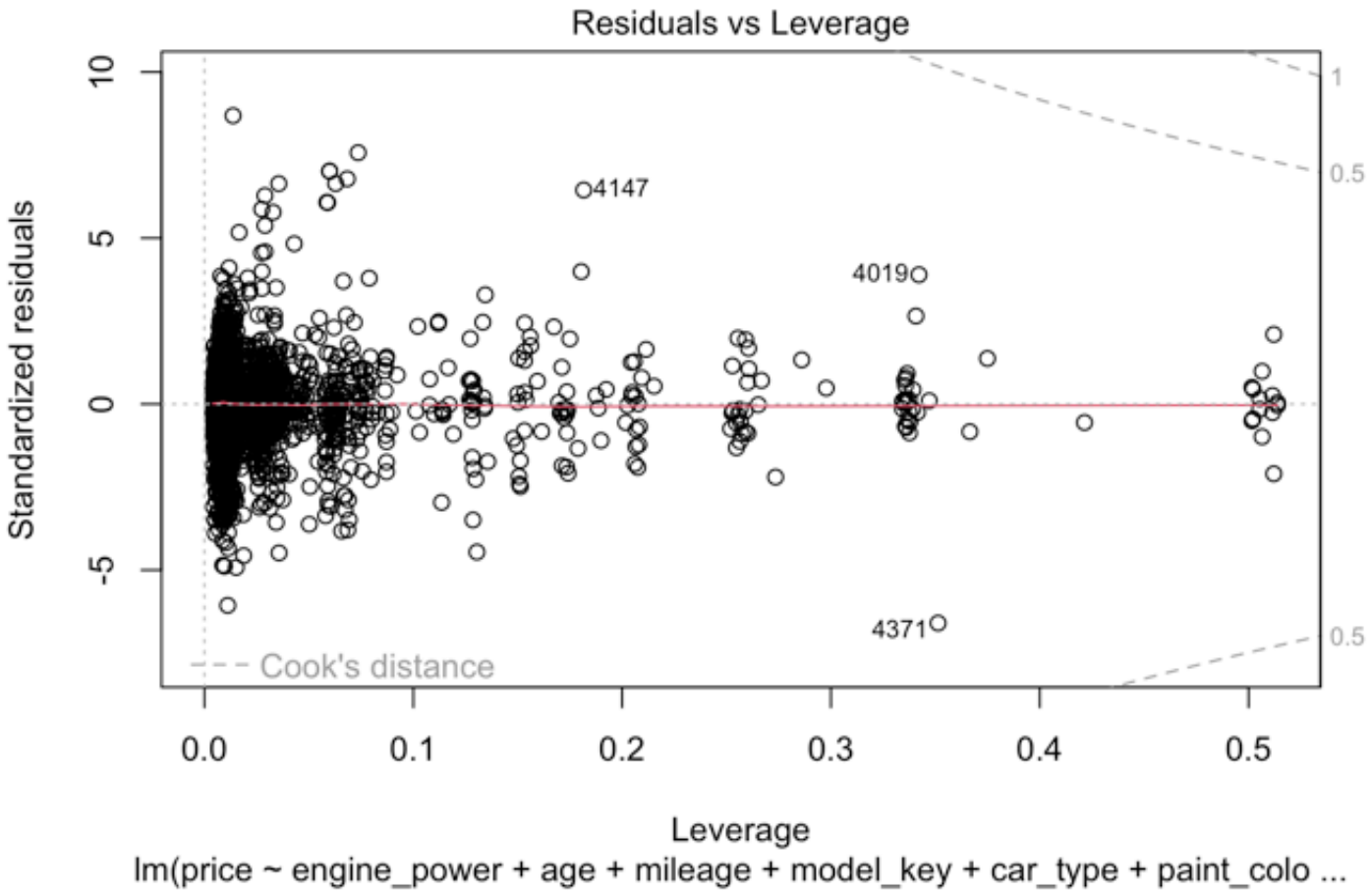
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3560 on 4725 degrees of freedom
## Multiple R-squared:  0.8317, Adjusted R-squared:  0.8278
## F-statistic: 210.4 on 111 and 4725 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
Residuals
Fitted values
lm(price ~ engine_power + age + mileage + model_key + car_type + paint_colo ...
3733
2679
3321
30000
10000
0
-20000
-20000
0
20000
40000
60000


Normal Q-Q
Standardized residuals
2679
3733
10
5
0
-5
-4
-2
0
2
4

Theoretical Quantiles
lm(price ~ engine_power + age + mileage + model_key + car_type + paint_colo ...





We observe that when plotting a naive MLR with all features we have a quadratic pattern in the plain and square-rooted standardized residuals and several points are marked as outliers and some point of high leverage.Before we spend time inquiring further, we must whittle down our model to the most important features.

According to the summary output, all features but Feature 2 ($p \approx 0.99$) and Feature 5 ($p \approx 0.2$) were significant. No colors had any significance. Month sold does interestingly have some significance in the event of it being September ($p \approx 0.03$) or August ($p \approx 0.095$).

It may be tempting to say that the car market is hotter during the later Summer season; however, it is hard to generalize this finding because the dataset was sampled from a specific auction house.

Whether a car is a convertible, hatchback, or estate seems informative, and a vast majority of the models contain informative information as well regarding price. We see that all but two models possess signficance codes indiciating $p \approx 0$.

With respect to fuel type, we see that if a vehicle has the fuel class hybrid-petrol, that provides significant information, but there are only 8 cases are of type hybrid-petrol, which give us limited power to detect the real effect.

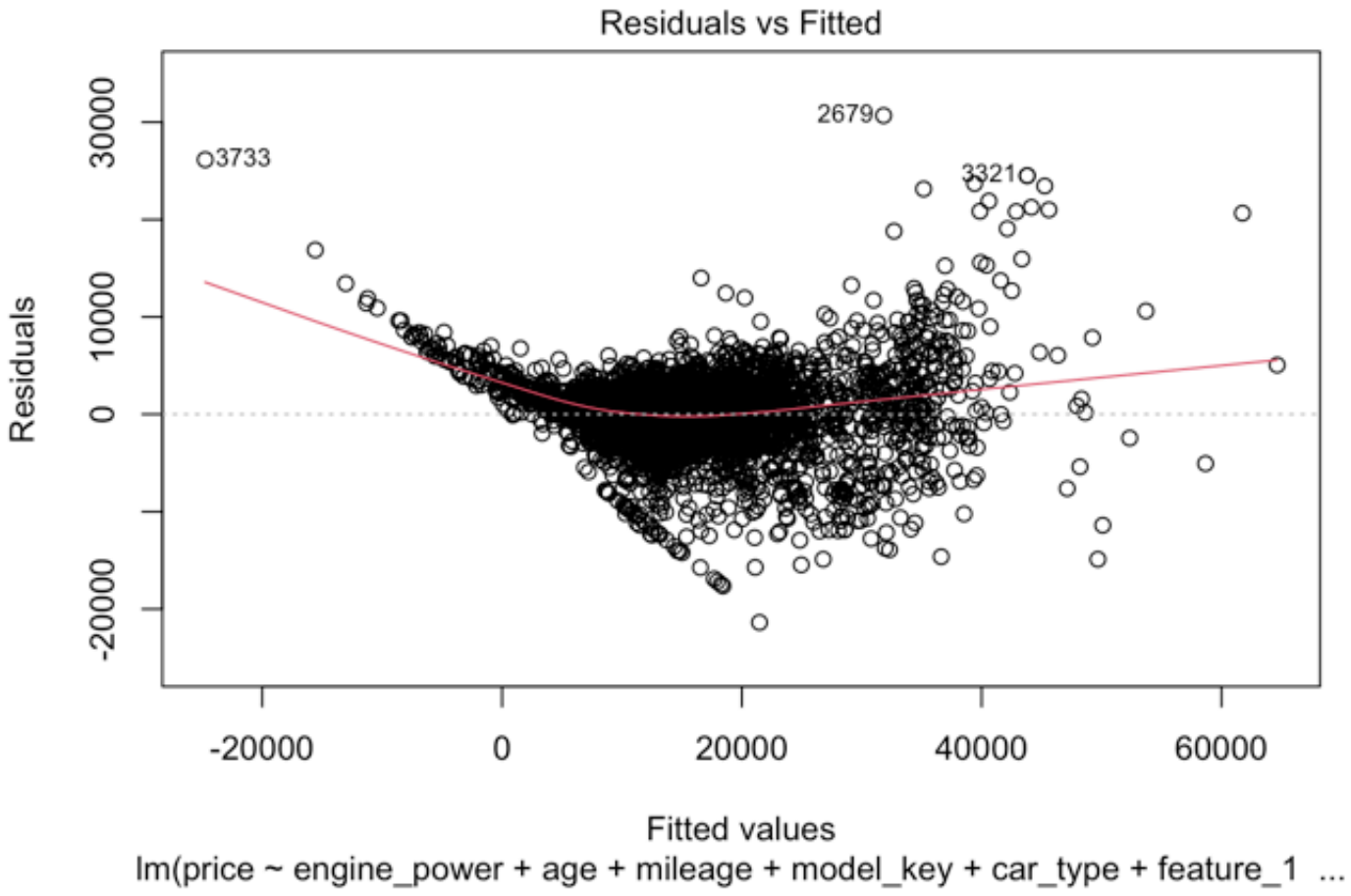
A new model could remove paint color, fuel, month sold, Feature 2, and Feature 5.

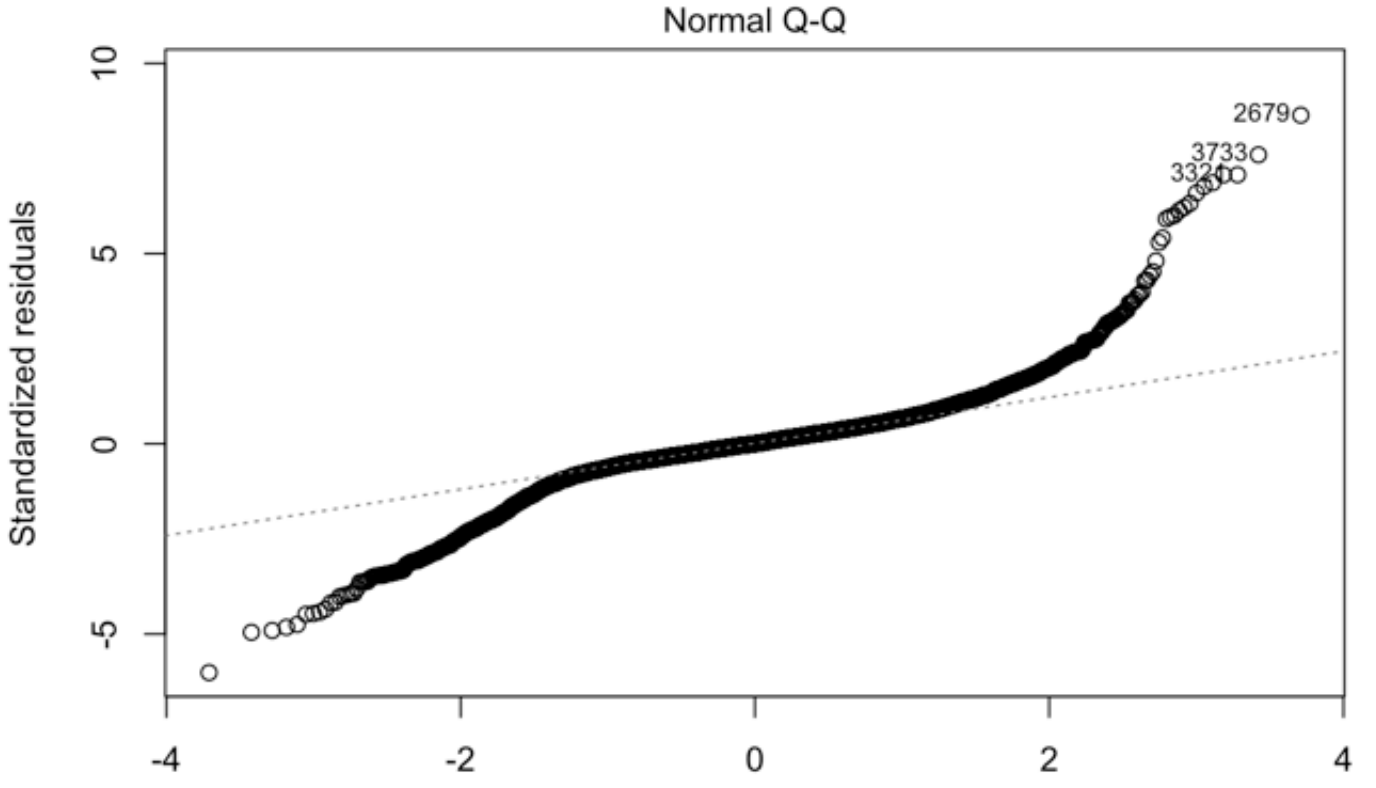
```
##
## Call:
```

```
## lm(formula = price ~ engine_power + age + mileage + model_key +
##     car_type + feature_1 + feature_3 + feature_4 + feature_6 +
##     feature_7 + feature_8, data = bmw_data_cleaned)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21373.6  -1398.9      0.4   1492.9  30678.9
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.603e+04  1.242e+03  12.906  < 2e-16 ***
## engine_power                5.171e+01  3.900e+00  13.259  < 2e-16 ***
## age                        -1.064e+03  2.645e+01 -40.211  < 2e-16 ***
## mileage                    -3.083e-02  1.065e-03 -28.962  < 2e-16 ***
## model_key116                1.945e+02  9.452e+02   0.206 0.836942
## model_key118                4.128e+02  9.831e+02   0.420 0.674578
## model_key120                3.393e+02  1.137e+03   0.299 0.765293
## model_key123               -2.517e+03  2.714e+03  -0.928 0.353707
## model_key125               -2.236e+03  1.879e+03  -1.190 0.234166
## model_key135                1.375e+03  2.120e+03   0.649 0.516639
## model_key214 Gran Tourer    9.423e+03  3.937e+03   2.394 0.016725 *
## model_key216               -3.482e+03  3.936e+03  -0.885 0.376345
## model_key216 Active Tourer  6.285e+03  2.644e+03   2.377 0.017478 *
## model_key216 Gran Tourer    7.451e+03  2.434e+03   3.061 0.002219 **
## model_key218                2.962e+03  1.497e+03   1.978 0.047953 *
## model_key218 Active Tourer  6.877e+03  1.912e+03   3.596 0.000326 ***
## model_key218 Gran Tourer    5.377e+03  2.041e+03   2.634 0.008463 **
## model_key220                2.686e+03  2.348e+03   1.144 0.252766
## model_key220 Active Tourer  2.386e+03  3.944e+03   0.605 0.545157
## model_key225               -8.084e+02  3.753e+03  -0.215 0.829441
## model_key225 Active Tourer  5.515e+03  3.954e+03   1.395 0.163165
## model_key316                2.155e+03  1.022e+03   2.108 0.035054 *
## model_key318                2.636e+03  1.015e+03   2.597 0.009422 **
## model_key318 Gran Turismo   3.596e+03  1.010e+03   3.559 0.000376 ***
## model_key320                2.081e+03  1.024e+03   2.032 0.042238 *
## model_key320 Gran Turismo   3.305e+03  1.057e+03   3.128 0.001770 **
## model_key325                7.697e+02  1.333e+03   0.577 0.563685
## model_key325 Gran Turismo   6.640e+03  2.044e+03   3.248 0.001169 **
## model_key328                1.584e+03  1.928e+03   0.821 0.411536
## model_key330                1.620e+03  1.255e+03   1.291 0.196930
## model_key330 Gran Turismo   4.190e+03  2.309e+03   1.815 0.069637 .
## model_key335                3.667e+03  1.779e+03   2.062 0.039303 *
## model_key335 Gran Turismo   3.728e+03  2.344e+03   1.590 0.111799
## model_key418 Gran Coupé     6.290e+03  1.580e+03   3.980 6.98e-05 ***
## model_key420                7.171e+03  1.222e+03   5.866 4.78e-09 ***
## model_key420 Gran Coupé     7.220e+03  1.148e+03   6.291 3.45e-10 ***
## model_key425                1.137e+04  2.767e+03   4.109 4.04e-05 ***
## model_key430                8.115e+03  2.781e+03   2.918 0.003541 **
## model_key430 Gran Coupé     6.582e+03  2.088e+03   3.153 0.001628 **
## model_key435                6.858e+03  2.187e+03   3.136 0.001721 **
```

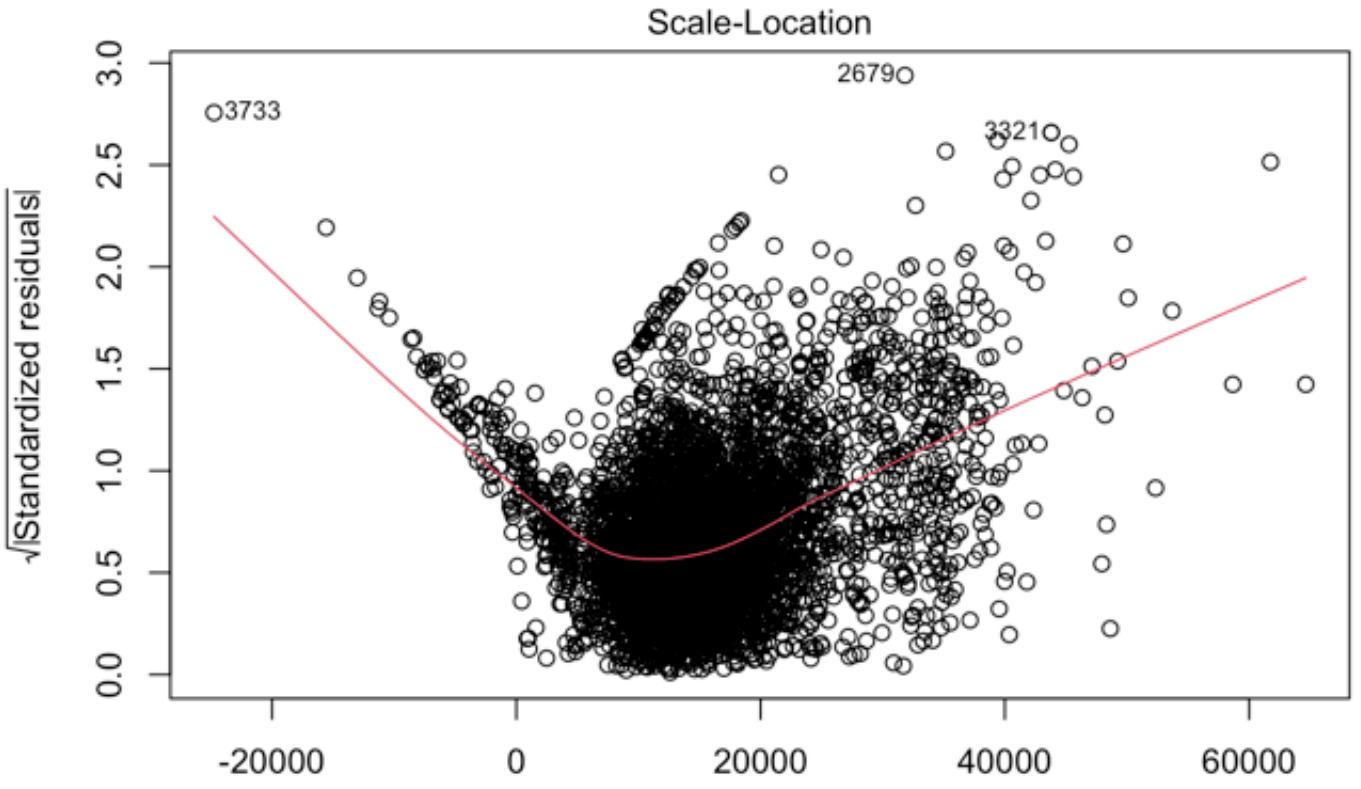
```
## model_key435 Gran Coupé     1.023e+04  1.833e+03   5.578 2.57e-08 ***
## model_key518                 5.223e+03  1.101e+03   4.744 2.16e-06 ***
## model_key520                 4.977e+03  1.034e+03   4.813 1.54e-06 ***
## model_key520 Gran Turismo    6.898e+03  1.224e+03   5.637 1.83e-08 ***
## model_key523                 6.711e+03  2.083e+03   3.221 0.001285 **
## model_key525                 4.617e+03  1.086e+03   4.251 2.16e-05 ***
## model_key528                 4.355e+03  1.820e+03   2.393 0.016763 *
## model_key530                 5.094e+03  1.133e+03   4.497 7.06e-06 ***
## model_key530 Gran Turismo    5.704e+03  1.352e+03   4.218 2.51e-05 ***
## model_key535                 5.445e+03  1.282e+03   4.246 2.21e-05 ***
## model_key535 Gran Turismo    3.901e+03  2.765e+03   1.411 0.158357
## model_key630                -6.286e+02  3.774e+03  -0.167 0.867710
## model_key635                 5.888e+03  3.782e+03   1.557 0.119552
## model_key640                 1.461e+04  1.788e+03   8.171 3.89e-16 ***
## model_key640 Gran Coupé      1.600e+04  1.411e+03  11.339  < 2e-16 ***
## model_key650                -1.157e+04  2.901e+03  -3.988 6.75e-05 ***
## model_key730                 9.407e+03  1.256e+03   7.490 8.14e-14 ***
## model_key735                 8.856e+02  3.764e+03   0.235 0.814013
## model_key740                 1.704e+04  1.446e+03  11.783  < 2e-16 ***
## model_key750                 1.120e+04  2.832e+03   3.954 7.80e-05 ***
## model_keyActiveHybrid 5      1.212e+03  3.757e+03   0.323 0.747036
## model_keyi3                  8.858e+03  1.731e+03   5.119 3.20e-07 ***
## model_keyM135                4.669e+03  3.749e+03   1.245 0.213112
## model_keyM235                4.644e+03  2.423e+03   1.916 0.055371 .
## model_keyM3                  1.636e+04  1.926e+03   8.496  < 2e-16 ***
## model_keyM4                  3.040e+04  2.898e+03  10.489  < 2e-16 ***
## model_keyM5                  1.378e+04  3.937e+03   3.501 0.000467 ***
## model_keyM550                1.117e+04  1.639e+03   6.816 1.05e-11 ***
## model_keyX1                  1.894e+03  1.907e+03   0.993 0.320558
## model_keyX3                  4.933e+03  1.909e+03   2.584 0.009788 **
## model_keyX4                  1.315e+04  1.996e+03   6.590 4.88e-11 ***
## model_keyX5                  1.388e+04  1.973e+03   7.035 2.28e-12 ***
## model_keyX5 M                1.749e+04  2.151e+03   8.131 5.38e-16 ***
## model_keyX5 M50              2.440e+04  2.924e+03   8.345  < 2e-16 ***
## model_keyX6                  1.621e+04  2.058e+03   7.874 4.22e-15 ***
## model_keyX6 M                2.612e+04  2.393e+03  10.913  < 2e-16 ***
## model_keyZ4                 -1.117e+02  1.888e+03  -0.059 0.952823
## car_typecoupe               -2.934e+03  7.251e+02  -4.047 5.28e-05 ***
## car_typeestate              -4.793e+03  6.874e+02  -6.972 3.54e-12 ***
## car_typehatchback           -3.341e+03  7.111e+02  -4.698 2.70e-06 ***
## car_typesedan               -2.770e+03  6.848e+02  -4.045 5.31e-05 ***
## car_typesubcompact          -2.495e+03  7.889e+02  -3.163 0.001574 **
## car_typesuv                 -3.633e+03  1.722e+03  -2.109 0.034966 *
## car_typevan                 -8.701e+03  1.437e+03  -6.054 1.52e-09 ***
## feature_1TRUE                5.758e+02  1.230e+02   4.681 2.93e-06 ***
## feature_3TRUE                6.887e+02  1.405e+02   4.901 9.85e-07 ***
## feature_4TRUE                1.375e+03  1.740e+02   7.906 3.29e-15 ***
## feature_6TRUE                1.310e+03  1.298e+02  10.099  < 2e-16 ***
## feature_7TRUE                1.015e+03  2.263e+02   4.483 7.52e-06 ***
## feature_8TRUE                1.298e+03  1.281e+02  10.133  < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3570 on 4747 degrees of freedom
## Multiple R-squared:   0.83,  Adjusted R-squared:  0.8268
## F-statistic: 260.4 on 89 and 4747 DF,  p-value: < 2.2e-16
```

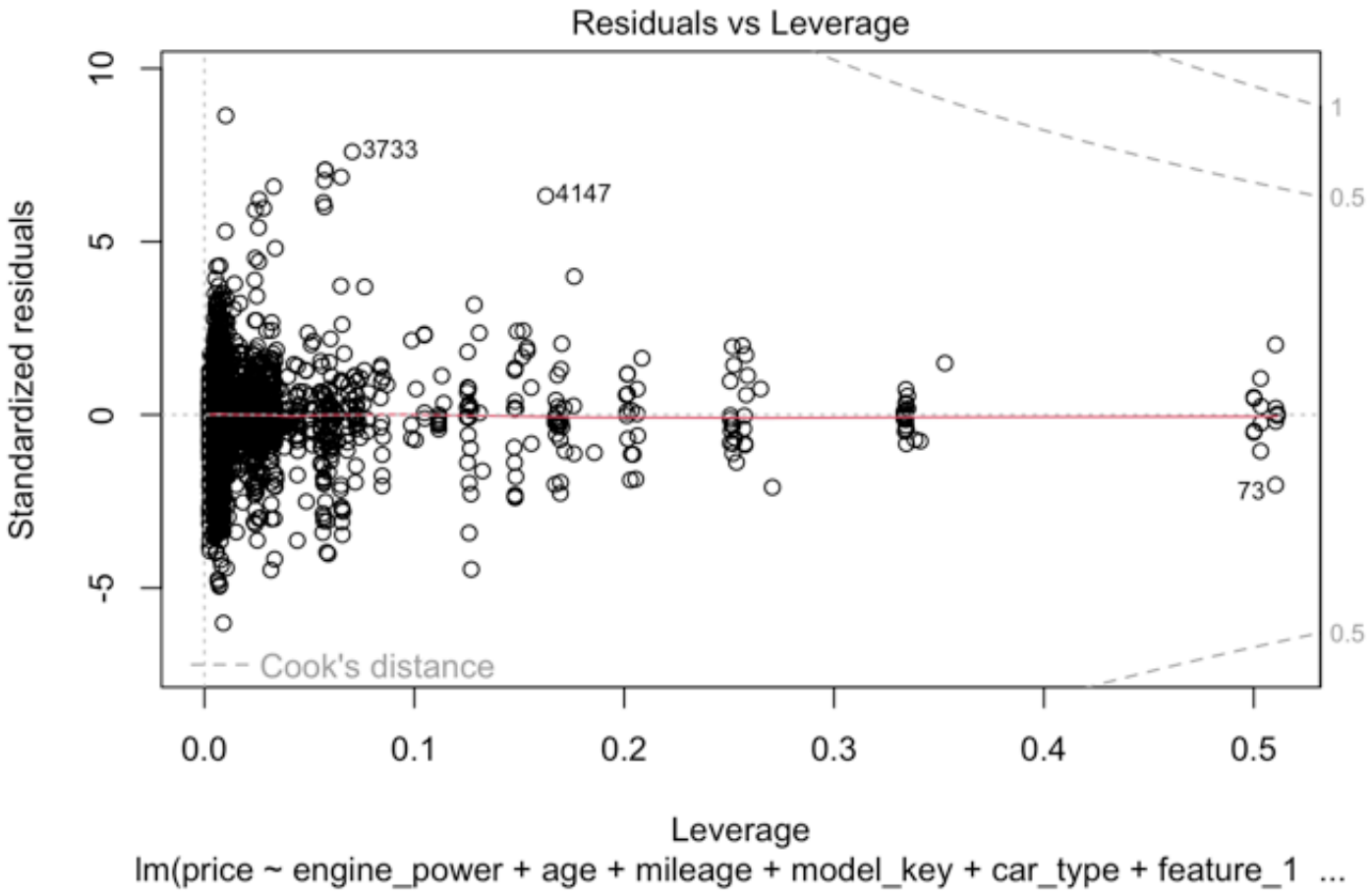

Residuals vs Fitted
Residuals
Fitted values
lm(price ~ engine_power + age + mileage + model_key + car_type + feature_1 ...
3733
2679
3321
30000
10000
0
-20000
-20000
0
20000
40000
60000


Normal Q-Q
Standardized residuals
2679
3733
3321
10
5
0
-5
-4
-2
0
2
4

Theoretical Quantiles
lm(price ~ engine_power + age + mileage + model_key + car_type + feature_1 ...



lm(price ~ engine_power + age + mileage + model_key + car_type + feature_1 ...

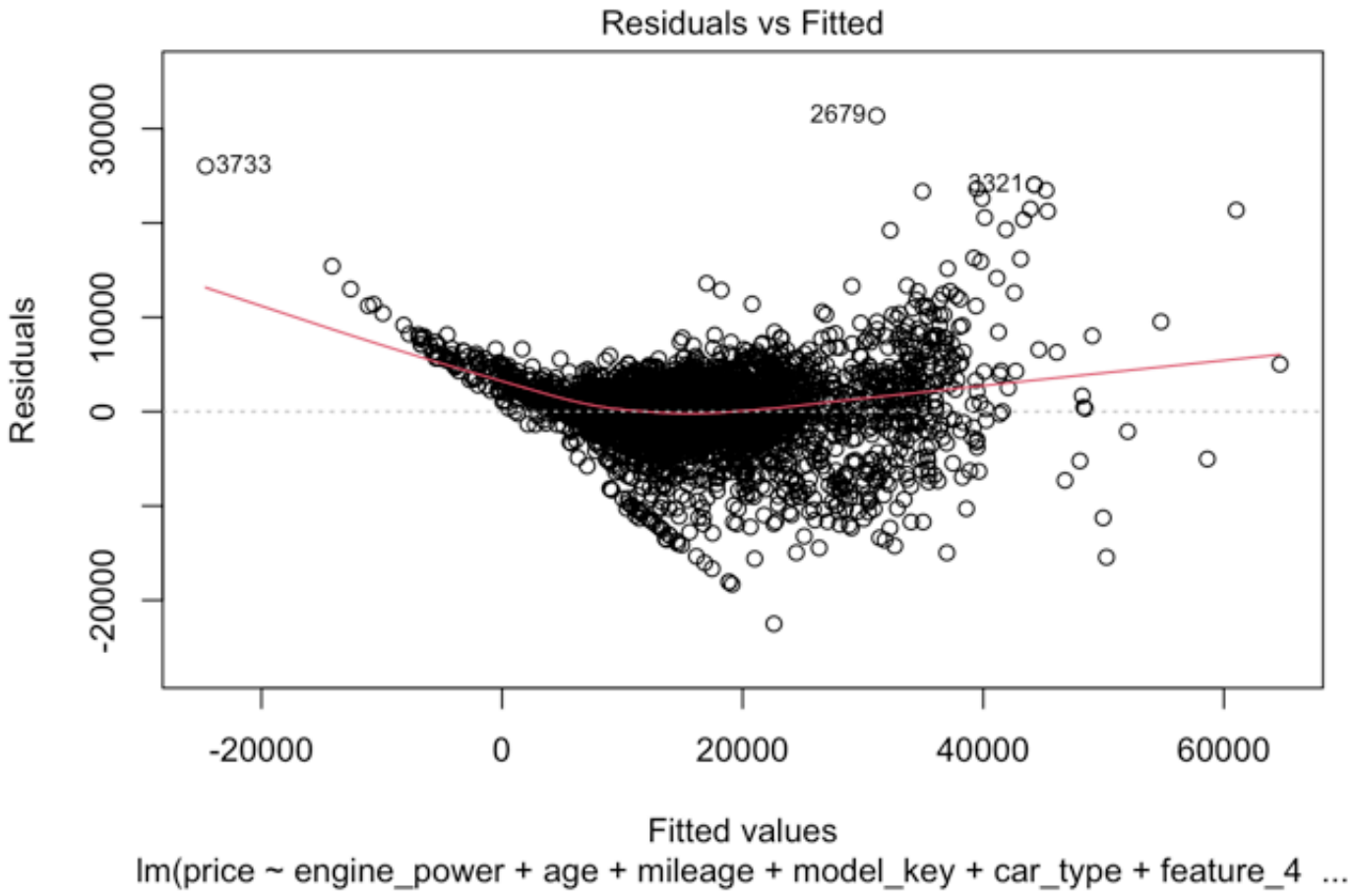


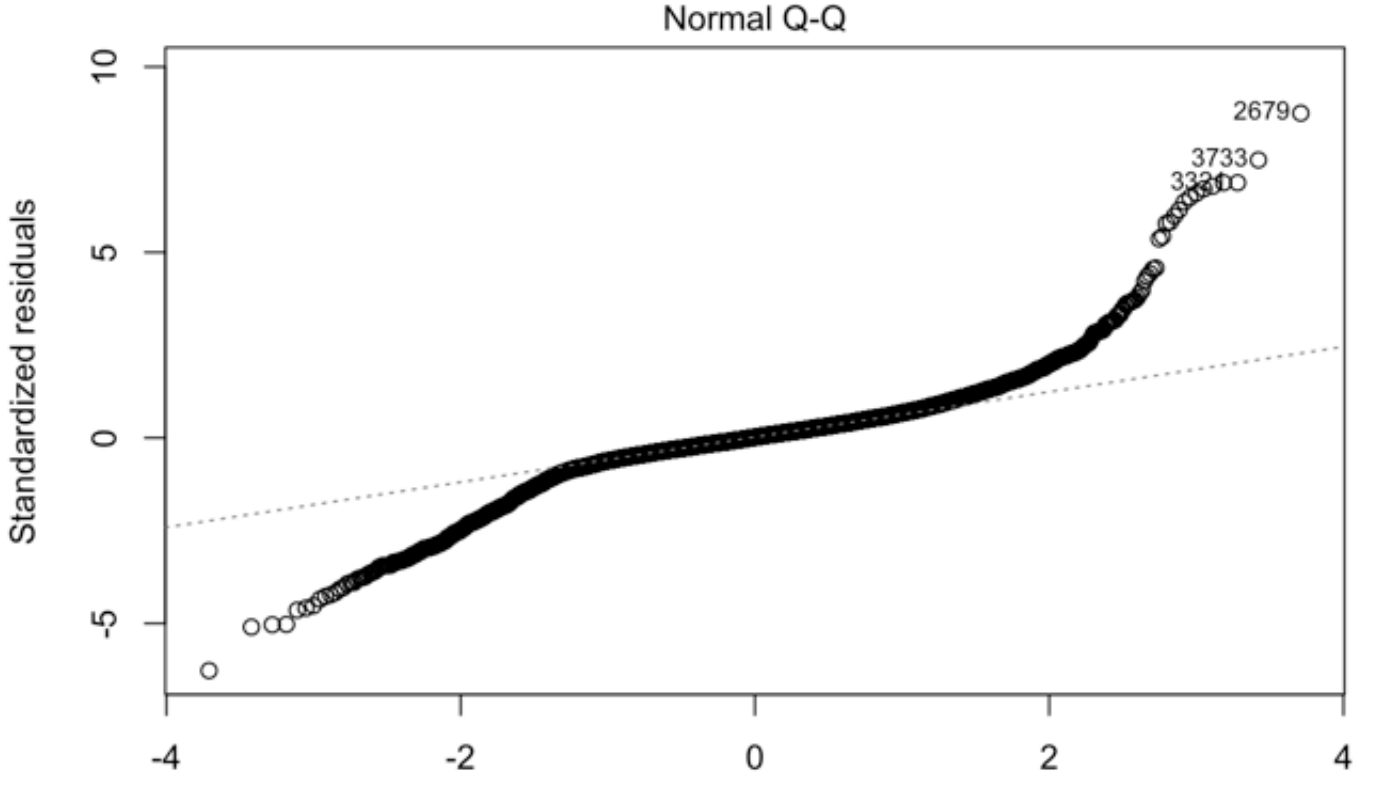
Our adjusted R^2 has remained roughly the same despite the removals, and our F-Statistic has continued to increase, from 210 to 260. We now keep only the three most statistically signficant features (4, 6, & 8).

```
##
## Call:
## lm(formula = price ~ engine_power + age + mileage + model_key +
##     car_type + feature_4 + feature_6 + feature_8, data = bmw_data_cleaned)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22503.8  -1371.8     52.1   1530.2  31364.4
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.707e+04  1.226e+03  13.918  < 2e-16 ***
## engine_power                5.154e+01  3.916e+00  13.164  < 2e-16 ***
## age                        -1.050e+03  2.658e+01 -39.491  < 2e-16 ***
## mileage                    -3.033e-02  1.065e-03 -28.472  < 2e-16 ***
## model_key116                2.668e+02  9.532e+02   0.280 0.779576
## model_key118                4.718e+02  9.915e+02   0.476 0.634171
## model_key120                3.657e+02  1.145e+03   0.319 0.749555
## model_key123               -1.867e+03  2.736e+03  -0.682 0.494994
## model_key125               -2.141e+03  1.896e+03  -1.129 0.258873
```

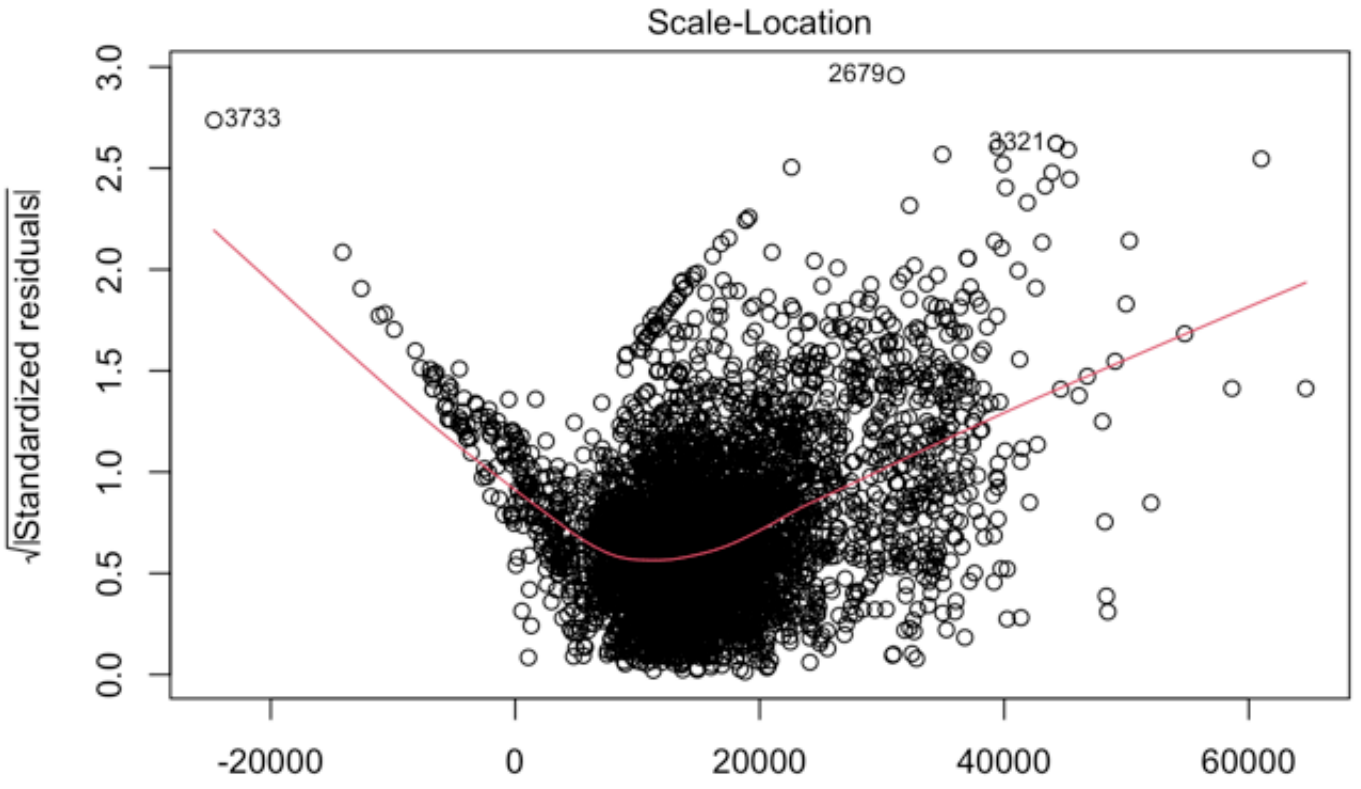
```
## model_key135                1.870e+03  2.138e+03   0.875 0.381829
## model_key214 Gran Tourer    1.061e+04  3.969e+03   2.672 0.007567 **
## model_key216               -3.879e+03  3.969e+03  -0.977 0.328468
## model_key216 Active Tourer  7.274e+03  2.665e+03   2.729 0.006368 **
## model_key216 Gran Tourer    8.251e+03  2.453e+03   3.363 0.000777 ***
## model_key218                2.948e+03  1.510e+03   1.952 0.051003 .
## model_key218 Active Tourer  7.594e+03  1.927e+03   3.941 8.22e-05 ***
## model_key218 Gran Tourer    6.191e+03  2.056e+03   3.011 0.002619 **
## model_key220                2.440e+03  2.368e+03   1.031 0.302807
## model_key220 Active Tourer  2.969e+03  3.977e+03   0.747 0.455359
## model_key225               -9.596e+02  3.786e+03  -0.253 0.799914
## model_key225 Active Tourer  5.719e+03  3.985e+03   1.435 0.151324
## model_key316                2.342e+03  1.030e+03   2.274 0.023023 *
## model_key318                2.915e+03  1.022e+03   2.853 0.004350 **
## model_key318 Gran Turismo   3.842e+03  1.018e+03   3.775 0.000162 ***
## model_key320                2.336e+03  1.031e+03   2.266 0.023509 *
## model_key320 Gran Turismo   3.592e+03  1.064e+03   3.376 0.000743 ***
## model_key325                1.145e+03  1.341e+03   0.853 0.393445
## model_key325 Gran Turismo   6.840e+03  2.062e+03   3.317 0.000917 ***
## model_key328                1.922e+03  1.944e+03   0.989 0.322779
## model_key330                2.039e+03  1.263e+03   1.615 0.106450
## model_key330 Gran Turismo   4.957e+03  2.326e+03   2.131 0.033140 *
## model_key335                4.310e+03  1.792e+03   2.405 0.016195 *
## model_key335 Gran Turismo   3.318e+03  2.364e+03   1.404 0.160451
## model_key418 Gran Coupé     6.886e+03  1.591e+03   4.329 1.53e-05 ***
## model_key420                7.336e+03  1.231e+03   5.960 2.71e-09 ***
## model_key420 Gran Coupé     7.739e+03  1.154e+03   6.705 2.25e-11 ***
## model_key425                1.104e+04  2.789e+03   3.959 7.63e-05 ***
## model_key430                8.185e+03  2.805e+03   2.918 0.003544 **
## model_key430 Gran Coupé     6.865e+03  2.102e+03   3.266 0.001100 **
## model_key435                7.281e+03  2.203e+03   3.305 0.000957 ***
## model_key435 Gran Coupé     1.067e+04  1.845e+03   5.786 7.67e-09 ***
## model_key518                5.893e+03  1.105e+03   5.335 9.99e-08 ***
## model_key520                5.564e+03  1.038e+03   5.361 8.67e-08 ***
## model_key520 Gran Turismo   7.774e+03  1.228e+03   6.330 2.68e-10 ***
## model_key523                6.834e+03  2.098e+03   3.257 0.001136 **
## model_key525                5.194e+03  1.091e+03   4.762 1.97e-06 ***
## model_key528                4.758e+03  1.835e+03   2.593 0.009532 **
## model_key530                5.717e+03  1.138e+03   5.026 5.20e-07 ***
## model_key530 Gran Turismo   6.392e+03  1.360e+03   4.700 2.67e-06 ***
## model_key535                6.161e+03  1.288e+03   4.782 1.79e-06 ***
## model_key535 Gran Turismo   4.835e+03  2.786e+03   1.735 0.082759 .
## model_key630                3.408e+02  3.805e+03   0.090 0.928646
## model_key635                7.091e+03  3.813e+03   1.860 0.062966 .
## model_key640                1.555e+04  1.799e+03   8.642  < 2e-16 ***
## model_key640 Gran Coupé     1.695e+04  1.418e+03  11.958  < 2e-16 ***
## model_key650               -1.114e+04  2.924e+03  -3.808 0.000142 ***
## model_key730                1.043e+04  1.260e+03   8.277  < 2e-16 ***
## model_key735                1.406e+03  3.795e+03   0.371 0.710977
## model_key740                1.804e+04  1.453e+03  12.417  < 2e-16 ***
```

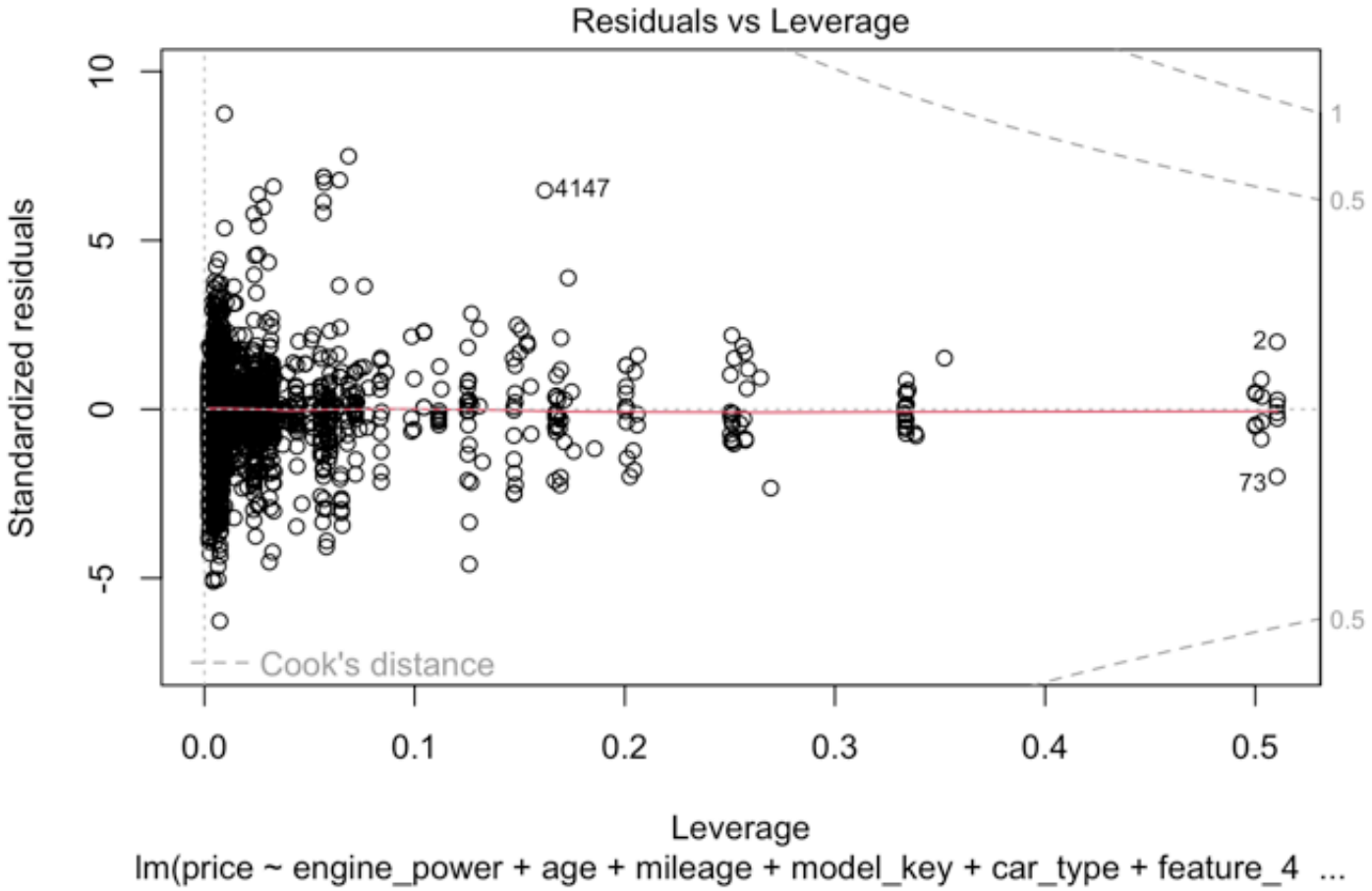
```
## model_key750                1.238e+04  2.853e+03   4.338 1.46e-05 ***
## model_keyActiveHybrid 5     1.726e+03  3.787e+03   0.456 0.648526
## model_keyi3                 9.302e+03  1.745e+03   5.331 1.02e-07 ***
## model_keyM135               5.322e+03  3.779e+03   1.408 0.159123
## model_keyM235               5.358e+03  2.443e+03   2.193 0.028355 *
## model_keyM3                 1.674e+04  1.938e+03   8.639  < 2e-16 ***
## model_keyM4                 3.082e+04  2.920e+03  10.554  < 2e-16 ***
## model_keyM5                 1.498e+04  3.969e+03   3.775 0.000162 ***
## model_keyM550               1.177e+04  1.650e+03   7.134 1.12e-12 ***
## model_keyX1                 2.124e+03  1.923e+03   1.104 0.269604
## model_keyX3                 5.301e+03  1.925e+03   2.754 0.005903 **
## model_keyX4                 1.367e+04  2.011e+03   6.795 1.22e-11 ***
## model_keyX5                 1.460e+04  1.988e+03   7.342 2.47e-13 ***
## model_keyX5 M               1.844e+04  2.167e+03   8.507  < 2e-16 ***
## model_keyX5 M50             2.516e+04  2.946e+03   8.542  < 2e-16 ***
## model_keyX6                 1.708e+04  2.074e+03   8.238 2.24e-16 ***
## model_keyX6 M               2.703e+04  2.412e+03  11.208  < 2e-16 ***
## model_keyZ4                 1.110e+02  1.903e+03   0.058 0.953486
## car_typecoupe              -3.046e+03  7.302e+02  -4.171 3.09e-05 ***
## car_typeestate             -5.015e+03  6.908e+02  -7.260 4.52e-13 ***
## car_typehatchback          -3.484e+03  7.151e+02  -4.872 1.14e-06 ***
## car_typesedan              -3.006e+03  6.884e+02  -4.366 1.29e-05 ***
## car_typesubcompact         -2.569e+03  7.935e+02  -3.238 0.001214 **
## car_typesuv                -3.958e+03  1.735e+03  -2.281 0.022574 *
## car_typevan                -9.380e+03  1.445e+03  -6.490 9.45e-11 ***
## feature_4TRUE               1.638e+03  1.718e+02   9.534  < 2e-16 ***
## feature_6TRUE               1.509e+03  1.291e+02  11.693  < 2e-16 ***
## feature_8TRUE               1.275e+03  1.285e+02   9.928  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3602 on 4750 degrees of freedom
## Multiple R-squared:  0.8269, Adjusted R-squared:  0.8237
## F-statistic: 263.8 on 86 and 4750 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
Residuals
Fitted values
lm(price ~ engine_power + age + mileage + model_key + car_type + feature_4 ...
2679
3733
3321
30000
10000
0
-20000
-20000
0
20000
40000
60000


Normal Q-Q
Standardized residuals
2679
3733
10
5
0
-5
-4
-2
0
2
4

Theoretical Quantiles
lm(price ~ engine_power + age + mileage + model_key + car_type + feature_4 ...





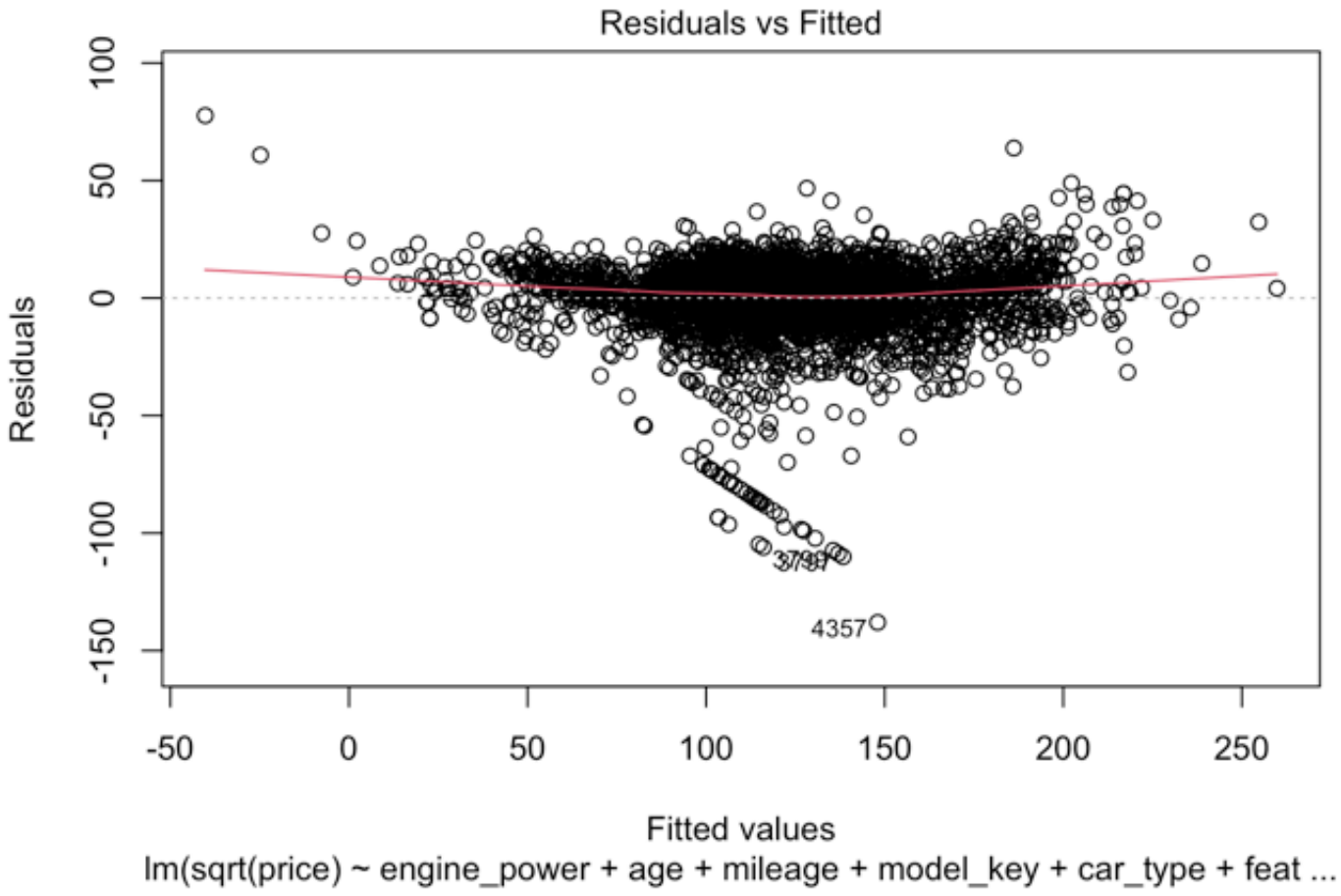
The adjusted R-squared remains consistent, while the F-statistics show a slight increase. We will retain the independent variables: engine power (horsepower), age (age), total mileage (mileage), model name (model_key), type (car_type), and specific features (feature_4, feature_6, feature_8).

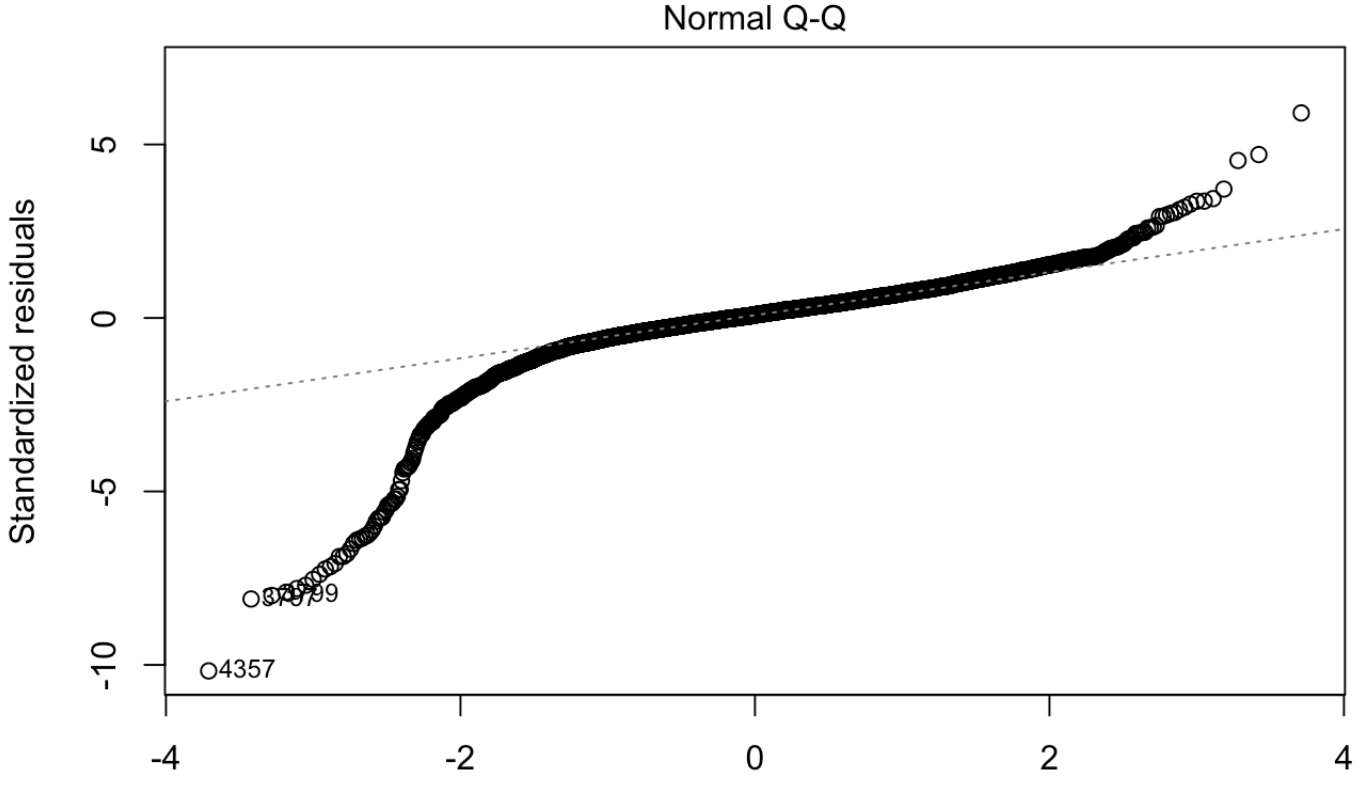
Upon inspecting the residual plot, a quadratic pattern emerges in both the residual vs. fitted and standard residual vs. fitted plots, suggesting the linear model may not fully capture the variables' relationship. This pattern implies a potential nonlinear relationship between the independent and dependent variables. As with the previously fitted simple linear regression, we consider applying a square root transformation to the price.

```
##
## Call:
## lm(formula = sqrt(price) ~ engine_power + age + mileage + model_key +
##     car_type + feature_4 + feature_6 + feature_8, data = bmw_data_cleaned)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -138.026   -4.544    1.152    6.739   77.666
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.388e+02  4.637e+00  29.934  < 2e-16 ***
## engine_power                1.498e-01  1.480e-02  10.117  < 2e-16 ***
## age                        -5.435e+00  1.005e-01 -54.084  < 2e-16 ***
```

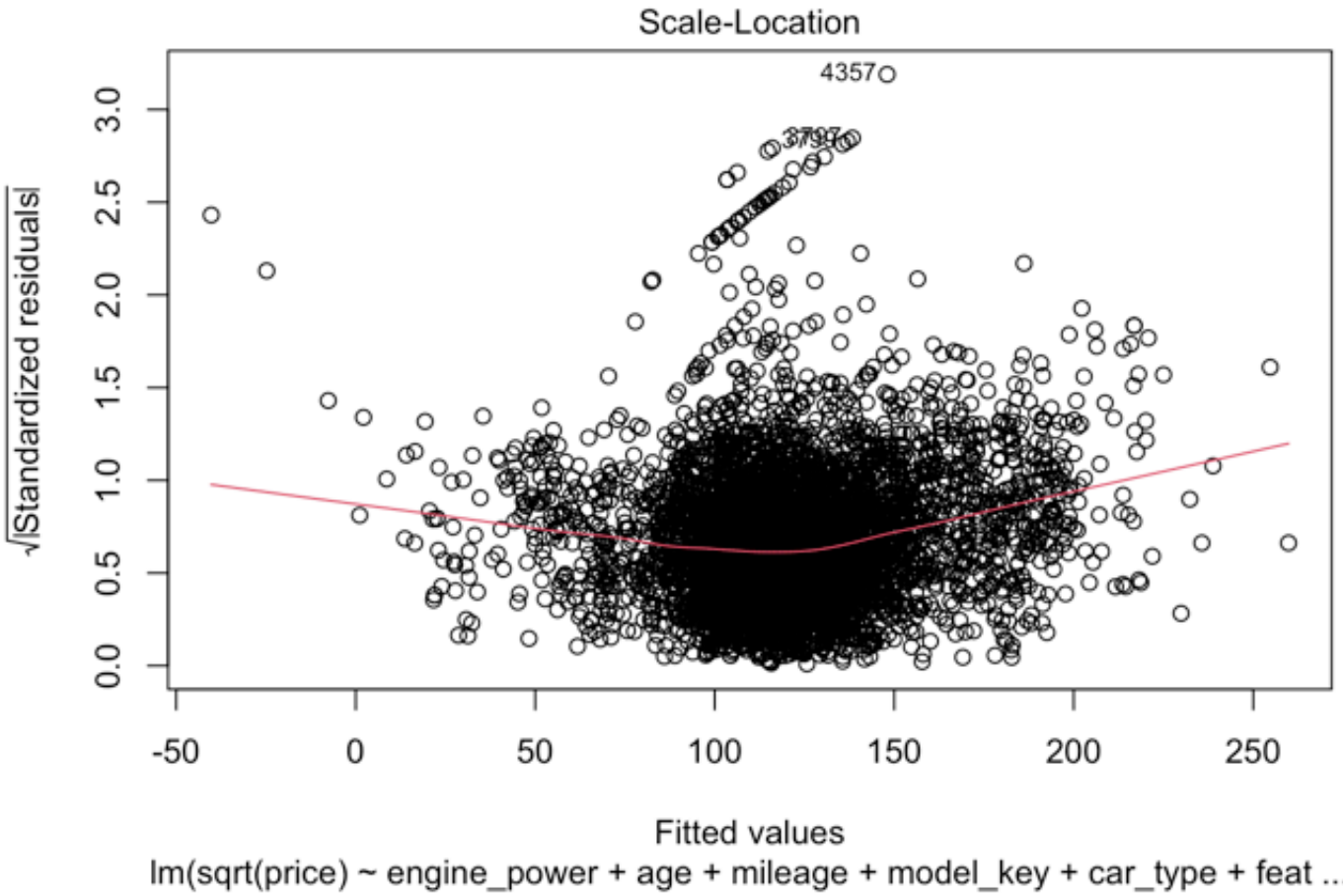
```
## mileage                    -1.088e-04  4.028e-06 -27.019  < 2e-16 ***
## model_key116                6.043e-01  3.604e+00   0.168 0.866835
## model_key118                1.397e+00  3.749e+00   0.373 0.709336
## model_key120                4.690e+00  4.331e+00   1.083 0.278850
## model_key123               -4.942e-01  1.034e+01  -0.048 0.961895
## model_key125               -3.324e+00  7.167e+00  -0.464 0.642848
## model_key135                1.424e+01  8.083e+00   1.762 0.078102 .
## model_key214 Gran Tourer    5.005e+01  1.501e+01   3.335 0.000858 ***
## model_key216               -2.013e+01  1.501e+01  -1.342 0.179815
## model_key216 Active Tourer  3.544e+01  1.008e+01   3.518 0.000439 ***
## model_key216 Gran Tourer    4.021e+01  9.276e+00   4.335 1.49e-05 ***
## model_key218                4.964e+00  5.709e+00   0.870 0.384617
## model_key218 Active Tourer  3.857e+01  7.285e+00   5.294 1.25e-07 ***
## model_key218 Gran Tourer    3.336e+01  7.774e+00   4.291 1.82e-05 ***
## model_key220                8.016e+00  8.952e+00   0.895 0.370576
## model_key220 Active Tourer  2.003e+01  1.504e+01   1.332 0.182866
## model_key225               -1.828e+00  1.431e+01  -0.128 0.898380
## model_key225 Active Tourer  3.514e+01  1.507e+01   2.332 0.019742 *
## model_key316                7.988e+00  3.895e+00   2.051 0.040344 *
## model_key318                1.187e+01  3.863e+00   3.073 0.002131 **
## model_key318 Gran Turismo   1.718e+01  3.849e+00   4.463 8.25e-06 ***
## model_key320                1.105e+01  3.898e+00   2.833 0.004627 **
## model_key320 Gran Turismo   1.739e+01  4.023e+00   4.322 1.58e-05 ***
## model_key325                1.064e+01  5.072e+00   2.098 0.035957 *
## model_key325 Gran Turismo   3.079e+01  7.796e+00   3.949 7.95e-05 ***
## model_key328                1.438e+01  7.350e+00   1.956 0.050474 .
## model_key330                1.322e+01  4.774e+00   2.769 0.005643 **
## model_key330 Gran Turismo   2.156e+01  8.795e+00   2.451 0.014288 *
## model_key335                2.232e+01  6.775e+00   3.294 0.000995 ***
## model_key335 Gran Turismo   1.750e+01  8.938e+00   1.958 0.050263 .
## model_key418 Gran Coupé     2.718e+01  6.014e+00   4.520 6.33e-06 ***
## model_key420                2.696e+01  4.654e+00   5.794 7.31e-09 ***
## model_key420 Gran Coupé     2.952e+01  4.364e+00   6.764 1.51e-11 ***
## model_key425                3.677e+01  1.055e+01   3.487 0.000493 ***
## model_key430                2.930e+01  1.061e+01   2.763 0.005757 **
## model_key430 Gran Coupé     2.531e+01  7.947e+00   3.184 0.001461 **
## model_key435                2.506e+01  8.329e+00   3.009 0.002638 **
## model_key435 Gran Coupé     3.517e+01  6.975e+00   5.043 4.76e-07 ***
## model_key518                2.461e+01  4.176e+00   5.892 4.08e-09 ***
## model_key520                2.542e+01  3.924e+00   6.479 1.02e-10 ***
## model_key520 Gran Turismo   3.342e+01  4.643e+00   7.197 7.12e-13 ***
## model_key523                2.309e+01  7.934e+00   2.910 0.003630 **
## model_key525                2.360e+01  4.123e+00   5.725 1.10e-08 ***
## model_key528                2.517e+01  6.936e+00   3.629 0.000288 ***
## model_key530                2.654e+01  4.301e+00   6.171 7.35e-10 ***
## model_key530 Gran Turismo   2.964e+01  5.141e+00   5.766 8.65e-09 ***
## model_key535                2.978e+01  4.871e+00   6.113 1.06e-09 ***
## model_key535 Gran Turismo   2.421e+01  1.053e+01   2.298 0.021577 *
## model_key630                1.256e+01  1.439e+01   0.873 0.382680
## model_key635                3.808e+01  1.442e+01   2.642 0.008279 **
```

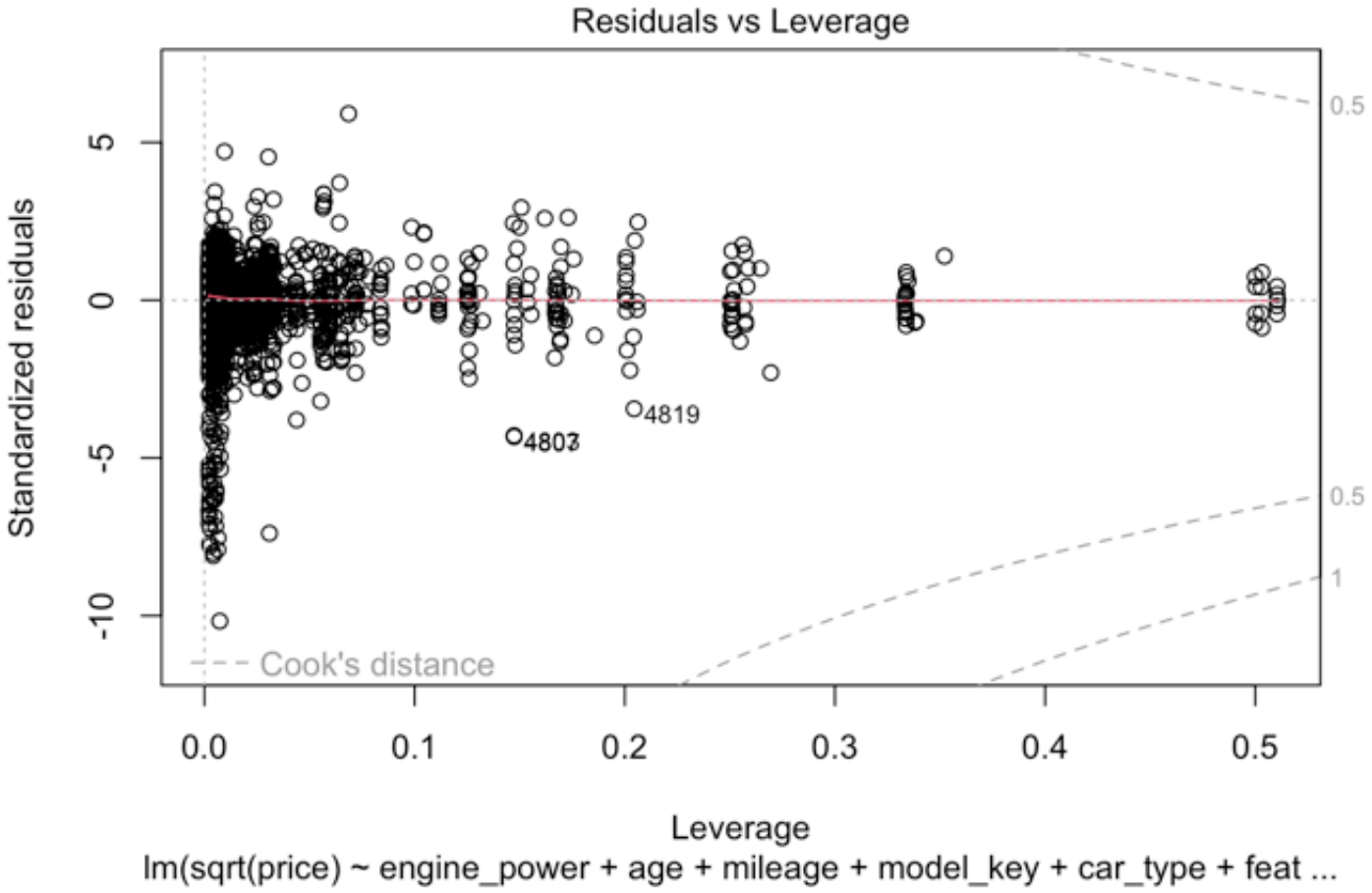
```
## model_key640                5.668e+01  6.802e+00   8.333  < 2e-16 ***
## model_key640 Gran Coupé     5.765e+01  5.361e+00  10.755  < 2e-16 ***
## model_key650               -2.987e+01  1.106e+01  -2.702 0.006922 **
## model_key730                4.330e+01  4.763e+00   9.092  < 2e-16 ***
## model_key735                1.348e+01  1.435e+01   0.940 0.347355
## model_key740                5.675e+01  5.492e+00  10.332  < 2e-16 ***
## model_key750                4.124e+01  1.079e+01   3.823 0.000134 ***
## model_keyActiveHybrid 5     1.616e+01  1.432e+01   1.129 0.259070
## model_keyi3                 3.589e+01  6.597e+00   5.440 5.58e-08 ***
## model_keyM135               2.620e+01  1.429e+01   1.834 0.066778 .
## model_keyM235               2.034e+01  9.237e+00   2.202 0.027702 *
## model_keyM3                 6.246e+01  7.327e+00   8.525  < 2e-16 ***
## model_keyM4                 7.316e+01  1.104e+01   6.625 3.85e-11 ***
## model_keyM5                 5.655e+01  1.500e+01   3.769 0.000166 ***
## model_keyM550               4.223e+01  6.238e+00   6.770 1.44e-11 ***
## model_keyX1                 1.194e+01  7.272e+00   1.642 0.100735
## model_keyX3                 2.565e+01  7.277e+00   3.525 0.000427 ***
## model_keyX4                 4.878e+01  7.604e+00   6.415 1.54e-10 ***
## model_keyX5                 5.427e+01  7.517e+00   7.219 6.08e-13 ***
## model_keyX5 M               6.318e+01  8.194e+00   7.710 1.52e-14 ***
## model_keyX5 M50             7.579e+01  1.114e+01   6.805 1.14e-11 ***
## model_keyX6                 6.179e+01  7.840e+00   7.882 3.97e-15 ***
## model_keyX6 M               7.688e+01  9.120e+00   8.430  < 2e-16 ***
## model_keyZ4                 6.773e+00  7.194e+00   0.941 0.346548
## car_typecoupe              -1.132e+01  2.761e+00  -4.100 4.20e-05 ***
## car_typeestate             -2.178e+01  2.612e+00  -8.339  < 2e-16 ***
## car_typehatchback          -1.542e+01  2.703e+00  -5.706 1.23e-08 ***
## car_typesedan              -1.286e+01  2.603e+00  -4.941 8.06e-07 ***
## car_typesubcompact         -1.306e+01  3.000e+00  -4.353 1.37e-05 ***
## car_typesuv                -1.872e+01  6.560e+00  -2.854 0.004338 **
## car_typevan                -4.606e+01  5.465e+00  -8.428  < 2e-16 ***
## feature_4TRUE               7.005e+00  6.494e-01  10.788  < 2e-16 ***
## feature_6TRUE               6.276e+00  4.880e-01  12.860  < 2e-16 ***
## feature_8TRUE               5.602e+00  4.857e-01  11.533  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.62 on 4750 degrees of freedom
## Multiple R-squared:  0.8309, Adjusted R-squared:  0.8278
## F-statistic: 271.4 on 86 and 4750 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
Residuals
100
50
0
-50
-100
-150
-50
0
50
100
150
200
250
4357
Fitted values
lm(sqrt(price) ~ engine_power + age + mileage + model_key + car_type + feat ...


Normal Q-Q
Standardized residuals
5
0
-5
-10
-4
-2
0
2
4
4357

Theoretical Quantiles
lm(sqrt(price) ~ engine_power + age + mileage + model_key + car_type + feat ...





Data Analysis Results

This output is from a multiple linear regression analysis where the square root of the car price (sqrt(price)) is modeled as a function of various predictors, including engine_power, age, mileage, different model_key categories, car_type, and features feature_4, feature_6, and feature_8, based on data from bmw_data_cleaned.

Intercept and coefficient

- The Intercept value of 200.6 suggests the average price($200.6^2 = 40240.36$) at baseline when all other predictors are zero.
- Engine Power For `engine_power` , the coefficient is 0.1498, indicating a positive relationship with the square root of the price. This means that as engine power increases, the square root of the price also increases, with each one-unit increase in engine power raising the square root of the price by approximately 0.1498 units.
- Age The coefficient for `age` is -5.435, showing a negative relationship with the square root of the price. As the age of the car increases by one year, the square root of the price decreases by about 5.435 units, reflecting the depreciation effect on the car's value.

- Mileage With a coefficient of -0.0001088 for `mileage`, there is a slight negative impact on the square root of the price. This implies that higher mileage slightly reduces the square root of the car's price, which could be due to wear and tear or perceived reduction in value.
- Model Key Categories Each `model_key` coefficient indicates the difference in the square root of the price for that model compared to a baseline model(M4). For example, `model_key750` has a coefficient of -31.92, suggesting that the average model's square root price is 31.92 lower compared to the average M4 model .
- Car Types The coefficients for `car_type` categories show the difference in square root price compared to the baseline car type(coupe). Convertibles (`car_typeconvertible`), for instance, have a positive coefficient of 11.32, indicating a higher square root price compared to the baseline, which suggests that convertibles are generally more expensive.
- Features Features like `feature_4` , `feature_6` , and `feature_8` have positive coefficients (7.005 for `feature_4` , for example), meaning that cars equipped with these features have a higher square root price. This indicates the added value or premium associated with these features in the car market.

Statistical Significance:

- Statistically significant predictors (p-value < 0.05) have a meaningful impact on the square root of the price. For example, age, mileage, and engine_power are significant, as well as several model_key and car_type categories.
- Model Fit: The Multiple R-squared value of 0.8309 indicates that approximately 83.09% of the variability in the square root of the car price is explained by the model, which is a strong level of explanatory power.

The Adjusted R-squared of 0.8278 adjusts this figure for the number of predictors in the model, confirming that the model fits well.

The F-statistic and its very small p-value (< 2.2e-16) suggest that the model is statistically significant; the predictors, as a whole, have a significant effect on the square root of the car price.

Residuals:

- The residuals' range and quartiles indicate how far off the model's predictions are from the actual values. Most residuals are within a reasonable range, suggesting no major issues with model fit across the data.
- The transformation of the dependent variable to its square root has notably enhanced the residual plots, indicating a better fit of the model. The standard residual versus fitted plot now shows a more flattened curve, suggesting increased constancy in the residuals across different fitted values.
- The quantile-quantile (QQ) plot has also shown improvement, particularly in the higher price range, signifying that the square root transformation of the price aligns more closely with the normal distribution assumption.

- However, an unusual grouping of data points is observed in the residual plot, with all residuals less than -50. Attempts to fit additional variables did not resolve this issue, implying that this specific data cluster warrants a more thorough investigation.