DSC 423: Data Analysis and Regression

Assignment 04: Model Building

Name: Adarsh Shankar

Student ID: 2117611

Honour Statement: “I have completed this work independently. The solutions given are entirely my own work." Your submission must be submitted as a PDF”.

Q1. CARPRICE. Find on the D2L a car price dataset. Use R to perform a regression analysis on the dataset Your submission should take the form of a technical report and should consider the following:

a. (10 pts.) Paste your final model into your submission (just the R output).

Console | Terminal | Background Jobs

R 4.2.1 · ~/Downloads/

```
> Carprice <- read.csv("/Users/adarsh/Downloads/CARPRICE.csv")
> View(Carprice)
> new <-Carprice[,-c(6,2,9,8,16,12,13,14,15)]
> model1 <- lm(price ~., data=new)
> summary(model1)

Call:
lm(formula = price ~ ., data = new)

Residuals:
    Min      1Q  Median      3Q     Max
-9575.0 -1884.6    -3.1  1613.0 12660.4

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -4.241e+04  1.852e+04  -2.290 0.023140 *
car_ID          -2.299e+01  4.507e+00  -5.100  8.2e-07 ***
fueltypegas     -8.622e+03  7.628e+03  -1.130 0.259764
aspirationturbo -3.218e+03  9.673e+02  -3.327 0.001053 **
doornumbertwo   -3.832e+02  6.235e+02  -0.615 0.539525
drivewheelfwd   -5.665e+02  1.354e+03  -0.418 0.676086
drivewheelrwd    1.415e+03  1.377e+03   1.028 0.305232
carlength        6.236e+01  5.008e+01   1.245 0.214543
carwidth         8.133e+02  2.413e+02   3.370 0.000911 ***
stroke          -1.660e+03  9.041e+02  -1.836 0.067975 .
compressionratio -2.783e+02  5.440e+02  -0.512 0.609554
horsepower       1.301e+02  1.228e+01  10.599  < 2e-16 ***
peakrpm         -6.221e-01  6.597e-01  -0.943 0.346857
citympg          2.631e+02  1.806e+02   1.457 0.146822
highwaympg      -2.221e+02  1.683e+02  -1.320 0.188488
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3431 on 190 degrees of freedom
Multiple R-squared:  0.8282,    Adjusted R-squared:  0.8155
F-statistic: 65.41 on 14 and 190 DF,  p-value: < 2.2e-16
```

b. (10 pts.) Describe the model building process through which you generated this model.

- I developed a number of models and included a large number of second order and interaction factors to gauge the variable's significance. For every model I tried, I did, however, get an adjusted R-squared value of 1.
  A variation of R-squared that takes into consideration variables in a regression model that are not significant is called adjusted R-squared. When compared to a model with more input variables, the adjusted R-squared is lower, indicating that the additional variables do not enhance the model.
  In comparison to a model with more input variables, a higher adjusted R-squared indicates that the additional input variables are improving the model. I too found an almost perfect correlation between price and square meters using the cor(CARPRICE) at first.
  This made it a perfect model in this instance. I attempted to drop square meters and construct a model in an effort to learn. At the conclusion of the task, I've attached the work.

c. (10 pts.) What significant second-order terms did you find, if any? Did you try all second-order terms? Did you look at scatter plots to determine which second-order terms to evaluate? Discuss the benefits and drawbacks of these two strategies.

- The correlation revealed a strong positive correlation between pricing and car width, curb weight, and horsepower. When these 3 terms were included in the second order model, the adj R2 jumped dramatically from 81.55% to 82.82%, and the p values also dropped significantly. Other phrases did not much improve the situation, and some even had detrimental impacts.

  The optimal second order terms were discovered to be car width, curb weight, and horsepower. Horsepower, curb weight, and car width had the strongest positive relationships in terms of the corelation. The relationships between other terms weren't very strong.

d. (10 pts.) What significant interaction terms did you find, if any? Did you try all combinations of interaction terms? Do you think that is an appropriate strategy? What happens to the number of interaction terms as the number of independent terms increases?

- "HorsePowerxcompressionratio" and "HorsePowerxstroke" in the model increased adjusted R-squared from 82.82% to 83.53%, improved F-test result, and reduced pvalues for coefficients, indicating enhanced model performance and increased significance of the interaction terms.
  The model did not show improvement with other attempted interaction terms such as "carwidth-curbweight," "carwidth-horsepower," and "curbweight-carwidth." The adjusted R-squared, F-test result, and p-values did not show significant changes, suggesting that these interaction terms may not have a strong impact on the model's performance.
  In a technical sense, horsepower and stroke are interconnected concepts in the context of the model. Considering the interaction between these two variables may have been meaningful as it could capture the combined effect of both variables on the model's performance, leading to improved results in terms of adjusted R-squared, F-test, and pvalues.

e. (10 pts.) Discuss your final model. Evaluate the t-tests, F-Test and adj-R2 accordingly. Do you think this is a “good” model? Explain.

- The second order terms and interaction terms' t tests pass muster, hence this model is valid.
  The regression model's independent variables are responsible for 83.53% of the price fluctuation.
  The null hypothesis can be rejected and the alternative accepted since at least one beta is not equal to 0.

  By that we can say that

  1. 104.4 on 10 and 194 DF for the F-test
  2. Adj R2: 83.53

f. Include your code an appendix.

Console | Terminal | Background Jobs

R 4.2.1 · ~/Downloads/

```
> new <-Carprice[,-c(1,6)]
> model2 <- lm(price ~., data=new)
> summary(model2)

Call:
lm(formula = price ~ ., data = new)

Residuals:
   Min     1Q Median     3Q    Max
 -2601      0      0      0   2262

Coefficients: (8 not defined because of singularities)
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -2.536e+04  2.928e+04  -0.866 0.392003
CarNamealfa-romero Quadrifoglio    7.409e+02  4.642e+03   0.160 0.874084
CarNamealfa-romero stelvio         3.005e+03  2.131e+03   1.410 0.167116
CarNameaudi 100 ls                 1.710e+03  3.941e+03   0.434 0.666921
CarNameaudi 100ls                  2.979e+03  4.608e+03   0.646 0.522158
CarNameaudi 4000                   5.626e+03  5.303e+03   1.061 0.295728
CarNameaudi 5000                   2.273e+03  4.862e+03   0.468 0.642958
CarNameaudi 5000s (diesel)        -5.426e+02  5.104e+03  -0.106 0.915933
CarNameaudi fox                    5.065e+03  4.699e+03   1.078 0.288277
CarNamebmw 320i                    4.122e+03  3.416e+03   1.206 0.235501
CarNamebmw x1                      1.304e+04  5.037e+03   2.588 0.013829 *
CarNamebmw x3                      1.371e+04  4.833e+03   2.836 0.007450 **
CarNamebmw x4                      1.490e+04  5.166e+03   2.884 0.006597 **
CarNamebmw x5                      2.408e+04  4.870e+03   4.944 1.78e-05 ***
CarNamebmw z4                      1.347e+04  5.024e+03   2.681 0.010993 *
CarNamebuick century               9.079e+03  4.996e+03   1.817 0.077524 .
CarNamebuick century luxus (sw)    6.954e+03  5.531e+03   1.257 0.216788
CarNamebuick century special       1.363e+04  4.566e+03   2.985 0.005073 **
CarNamebuick electra 225 custom    5.860e+03  5.180e+03   1.131 0.265449
CarNamebuick opel isuzu deluxe     7.890e+03  4.330e+03   1.822 0.076755 .
CarNamebuick regal sport coupe (turbo)  2.092e+04  4.175e+03   5.012 1.45e-05 ***
CarNamebuick skyhawk               8.868e+03  5.161e+03   1.718 0.094322 .
CarNamebuick skylark               9.809e+03  3.798e+03   2.583 0.014015 *
CarNamechevrolet impala            2.611e+03  4.087e+03   0.639 0.526989
CarNamechevrolet monte carlo      -3.446e+03  3.625e+03  -0.950 0.348259
CarNamechevrolet vega 2300        -3.666e+03  3.624e+03  -1.012 0.318418
CarNamedodge challenger se        -3.497e+03  3.463e+03  -1.010 0.319406
CarNamedodge colt (sw)            -4.257e+03  3.512e+03  -1.212 0.233420
CarNamedodge colt hardtop         -5.174e+03  3.512e+03  -1.473 0.149429
CarNamedodge coronet custom       -5.612e+03  3.686e+03  -1.523 0.136613
CarNamedodge coronet custom (sw)  -4.844e+03  3.510e+03  -1.380 0.176101
CarNamedodge d200                 -5.465e+03  3.655e+03  -1.495 0.143616
CarNamedodge dart custom          -3.050e+03  4.252e+03  -0.717 0.477748
```

```
CarNamedodge dart custom             -3.050e+03  4.252e+03  -0.717 0.477748
CarNamedodge monaco (sw)             -5.404e+03  3.499e+03  -1.544 0.131287
CarNamedodge rampage                 -4.582e+03  3.566e+03  -1.285 0.207023
CarNamehonda accord                  -2.826e+03  3.662e+03  -0.772 0.445190
CarNamehonda accord cvcc             -6.055e+03  3.837e+03  -1.578 0.123289
CarNamehonda accord lx               -4.127e+03  3.856e+03  -1.070 0.291606
CarNamehonda civic                   -2.458e+03  3.445e+03  -0.713 0.480256
CarNamehonda civic (auto)            -3.198e+03  3.531e+03  -0.906 0.371175
CarNamehonda civic 1300              -4.153e+03  3.706e+03  -1.121 0.269838
CarNamehonda civic 1500 gl           -3.926e+03  4.337e+03  -0.905 0.371436
CarNamehonda civic cvcc              -5.007e+03  3.594e+03  -1.393 0.172058
CarNamehonda prelude                 -3.362e+03  3.684e+03  -0.913 0.367546
CarNameisuzu D-Max                   -2.081e+03  3.152e+03  -0.660 0.513338
CarNameisuzu D-Max V-Cross           -1.909e+03  3.650e+03  -0.523 0.604179
CarNameisuzu MU-X                    -1.063e+03  3.289e+03  -0.323 0.748349
CarNamejaguar xf                      6.860e+03  4.053e+03   1.692 0.099191 .
CarNamejaguar xj                      3.560e+03  4.053e+03   0.878 0.385600
CarNamejaguar xk                      8.505e+03  6.427e+03   1.323 0.194102
CarNamemaxda glc deluxe              -1.377e+03  3.424e+03  -0.402 0.689918
CarNamemaxda rx3                     -1.571e+03  3.347e+03  -0.469 0.641555
CarNamemazda 626                     -6.611e+02  3.083e+03  -0.214 0.831431
CarNamemazda glc                      1.555e+03  3.350e+03   0.464 0.645250
CarNamemazda glc 4                    3.313e+03  3.856e+03   0.859 0.395983
CarNamemazda glc custom              -2.819e+02  3.275e+03  -0.086 0.931885
CarNamemazda glc custom l            -2.635e+03  3.410e+03  -0.773 0.444794
CarNamemazda glc deluxe               4.106e+01  3.294e+03   0.012 0.990125
CarNamemazda rx-4                    -6.509e+02  3.219e+03  -0.202 0.840875
CarNamemazda rx-7 gs                  2.930e+03  3.358e+03   0.872 0.388815
CarNamemazda rx2 coupe               -7.300e+02  3.423e+03  -0.213 0.832340
CarNamemercury cougar                -1.546e+03  4.201e+03  -0.368 0.715090
CarNamemitsubishi g4                 -4.489e+03  3.188e+03  -1.408 0.167657
CarNamemitsubishi lancer             -4.799e+03  3.448e+03  -1.392 0.172576
CarNamemitsubishi mirage             -5.610e+03  3.538e+03  -1.586 0.121568
CarNamemitsubishi mirage g4          -5.177e+03  3.255e+03  -1.591 0.120404
CarNamemitsubishi montero            -5.493e+03  3.156e+03  -1.741 0.090292 .
CarNamemitsubishi outlander          -5.506e+03  3.241e+03  -1.699 0.097994 .
CarNamemitsubishi pajero             -4.717e+03  3.167e+03  -1.489 0.145160
CarNamenissan clipper                -1.253e+03  3.387e+03  -0.370 0.713554
CarNamenissan dayz                   -3.930e+03  4.649e+03  -0.845 0.403497
CarNamenissan fuga                   -4.172e+03  4.913e+03  -0.849 0.401370
CarNamenissan gt-r                   -9.421e+02  3.700e+03  -0.255 0.800471
CarNamenissan juke                   -9.006e+01  3.485e+03  -0.026 0.979529
CarNamenissan kicks                  -1.293e+03  4.179e+03  -0.310 0.758696
CarNamenissan latio                  -2.268e+02  3.402e+03  -0.067 0.947222
CarNamenissan leaf                    6.495e+02  3.614e+03   0.180 0.858378
```

```
CarNamenissan note                      3.076e+02  3.480e+03   0.088 0.930052
CarNamenissan nv200                    -1.710e+03  3.508e+03  -0.487 0.628915
CarNamenissan otti                     -3.241e+03  4.729e+03  -0.685 0.497534
CarNamenissan rogue                    -1.097e+03  3.378e+03  -0.325 0.747296
CarNamenissan teana                    -1.910e+03  3.866e+03  -0.494 0.624306
CarNamenissan titan                    -2.047e+02  3.479e+03  -0.059 0.953421
CarNameNissan versa                    -4.936e+02  3.624e+03  -0.136 0.892429
CarNamepeugeot 304                     -8.562e+03  3.426e+03  -2.499 0.017152 *
CarNamepeugeot 504                     -6.657e+03  3.041e+03  -2.189 0.035162 *
CarNamepeugeot 504 (sw)                -9.587e+03  3.770e+03  -2.543 0.015433 *
CarNamepeugeot 505s turbo diesel       -5.687e+03  3.720e+03  -1.529 0.135061
CarNamepeugeot 604sl                   -6.253e+03  3.221e+03  -1.941 0.060111 .
CarNameplymouth cricket                -5.465e+03  3.655e+03  -1.495 0.143616
CarNameplymouth duster                 -4.913e+03  3.366e+03  -1.460 0.153081
CarNameplymouth fury gran sedan        -4.302e+03  3.559e+03  -1.209 0.234601
CarNameplymouth fury iii               -5.219e+03  3.346e+03  -1.560 0.127530
CarNameplymouth satellite custom (sw)  -3.079e+03  3.445e+03  -0.894 0.377326
CarNameplymouth valiant                -3.030e+03  4.251e+03  -0.713 0.480550
CarNameporcshce panamera                4.298e+03  5.939e+03   0.724 0.473884
CarNameporsche boxter                   8.332e+03  5.917e+03   1.408 0.167679
CarNameporsche cayenne                  5.798e+03  5.543e+03   1.046 0.302519
CarNameporsche macan                    4.254e+03  3.537e+03   1.203 0.236967
CarNamerenault 12tl                    -2.628e+03  3.677e+03  -0.715 0.479416
CarNamerenault 5 gtl                   -2.962e+03  3.414e+03  -0.867 0.391433
CarNamesaab 99e                         9.489e+02  3.655e+03   0.260 0.796656
CarNamesaab 99gle                       2.217e+03  3.650e+03   0.607 0.547385
CarNamesaab 99le                        5.755e+02  3.964e+03   0.145 0.885386
CarNamesubaru                          -4.405e+03  2.595e+03  -1.698 0.098208 .
CarNamesubaru baja                     -4.112e+03  2.585e+03  -1.591 0.120366
CarNamesubaru brz                      -2.475e+03  2.769e+03  -0.894 0.377468
CarNamesubaru dl                       -2.915e+03  2.806e+03  -1.039 0.305848
CarNamesubaru r1                       -2.642e+03  3.205e+03  -0.824 0.415188
CarNamesubaru r2                       -3.623e+03  3.292e+03  -1.101 0.278328
CarNamesubaru trezia                   -4.494e+03  2.668e+03  -1.684 0.100742
CarNamesubaru tribeca                  -4.438e+03  2.582e+03  -1.718 0.094300 .
CarNametoyota carina                   -7.133e+03  4.734e+03  -1.507 0.140650
CarNametoyota celica gt                -2.065e+03  3.783e+03  -0.546 0.588488
CarNametoyota celica gt liftback       -5.622e+03  3.124e+03  -1.800 0.080264 .
CarNametoyota corolla                  -1.780e+03  3.009e+03  -0.592 0.557772
CarNametoyota corolla 1200             -2.496e+03  3.134e+03  -0.797 0.430897
CarNametoyota corolla 1600 (sw)         6.754e+02  4.254e+03   0.159 0.874733
CarNametoyota corolla liftback         -1.670e+02  3.005e+03  -0.056 0.955998
CarNametoyota corolla tercel           -5.753e+03  3.132e+03  -1.837 0.074457 .
CarNametoyota corona                   -1.701e+03  2.908e+03  -0.585 0.562186
CarNametoyota corona hardtop           -8.950e+02  4.201e+03  -0.213 0.832498
CarNametoyota corona liftback          -3.223e+03  3.384e+03  -0.952 0.347232
```

```
CarNamevolvo 244dl                   1.013e+03  3.884e+03   0.261 0.795643
CarNamevolvo 245                    -6.447e+01  4.221e+03  -0.015 0.987900
CarNamevolvo 246                     5.019e+03  5.377e+03   0.933 0.356811
CarNamevolvo 264gl                   1.377e+03  3.970e+03   0.347 0.730783
CarNamevolvo diesel                  1.557e+02  4.288e+03   0.036 0.971229
CarNamevw dasher                     4.266e+01  3.640e+03   0.012 0.990714
CarNamevw rabbit                    -1.497e+03  3.334e+03  -0.449 0.656143
fueltypegas                         -8.764e+03  1.158e+04  -0.757 0.453942
aspirationturbo                      2.741e+02  1.096e+03   0.250 0.803951
doornumbertwo                       -7.142e+02  5.947e+02  -1.201 0.237578
drivewheelfwd                        2.502e+02  1.523e+03   0.164 0.870451
drivewheelrwd                        2.555e+01  1.850e+03   0.014 0.989059
enginelocationrear                   1.531e+04  3.191e+03   4.798 2.78e-05 ***
wheelbase                            3.028e+02  1.500e+02   2.019 0.051000 .
carlength                           -2.014e+02  7.170e+01  -2.809 0.007976 **
carwidth                             6.560e+02  3.540e+02   1.853 0.072106 .
carheight                           -4.035e+02  3.377e+02  -1.195 0.239988
curbweight                           1.060e+01  2.813e+00   3.767 0.000592 ***
enginetypedohcv                             NA         NA      NA       NA
enginetypel                                 NA         NA      NA       NA
enginetypeohc                       -2.226e+03  2.049e+03  -1.086 0.284615
enginetypeohcf                              NA         NA      NA       NA
enginetypeohcv                       9.958e+02  2.411e+03   0.413 0.681996
enginetyperotor                     -1.379e+03  2.776e+03  -0.497 0.622422
cylindernumberfive                          NA         NA      NA       NA
cylindernumberfour                   4.173e+03  2.468e+03   1.691 0.099489 .
cylindernumbersix                           NA         NA      NA       NA
cylindernumberthree                         NA         NA      NA       NA
cylindernumbertwelve                        NA         NA      NA       NA
cylindernumbertwo                           NA         NA      NA       NA
boreratio                           -2.006e+03  2.188e+03  -0.917 0.365323
stroke                              -9.028e-01  1.366e+03  -0.001 0.999476
compressionratio                    -6.822e+02  8.644e+02  -0.789 0.435138
horsepower                           6.277e+00  2.886e+01   0.217 0.829055
peakrpm                              2.390e+00  9.964e-01   2.399 0.021745 *
citympg                              5.388e+00  2.139e+02   0.025 0.980038
highwaympg                           8.492e+01  1.853e+02   0.458 0.649414
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1507 on 36 degrees of freedom
Multiple R-squared:  0.9937,    Adjusted R-squared:  0.9644
F-statistic: 33.91 on 168 and 36 DF,  p-value: < 2.2e-16

> 
```