Veera mukesh aripaka - U97302307

2022-10-11

#Veera mukesh aripaka - U97302307  
  
#Preprocessing:  
  
library(readxl)  
M5 <- read\_excel("C:/Users/mukes/Downloads/6304 Module 5 Assignment Data.xlsx")  
  
colnames(M5)=tolower(make.names(colnames(M5)))  
attach(M5)  
  
M5$year = as.integer(M5$year)  
attach(M5)

## The following objects are masked from M5 (pos = 3):  
##   
## condition, cylinders, fuel, make, model, odometer, paint.color,  
## price, region, year

M5subs = subset(M5,M5$make == "cadillac")  
M5subs = subset(M5subs,M5subs$year >= 2006 & M5subs$year <= 2011)  
M5subs1 = subset(M5subs,M5subs$condition == "excellent" | M5subs$condition=="good")  
M5subs2 = subset(M5subs1,M5subs1$cylinders == 6 | M5subs1$cylinders==8)  
M5subs3 = subset(M5subs2,M5subs2$paint.color != "black")  
M5subs3 = subset(M5subs3,M5subs3$paint.color != "custom")  
  
set.seed(97302307)  
s90 = M5subs3[sample(1:nrow(M5subs3),90),]  
  
#Analysis  
  
#⦁ Use the str() command to show your primary data set meets all the conditions listed Preprocessing steps 2 and 3 above  
  
str(s90)

## tibble [90 × 10] (S3: tbl\_df/tbl/data.frame)  
## $ region : chr [1:90] "western IL" "oklahoma city" "southwest MN" "colorado springs" ...  
## $ price : num [1:90] 6900 14500 5995 17999 26000 ...  
## $ year : int [1:90] 2006 2007 2007 2007 2011 2010 2006 2006 2006 2011 ...  
## $ make : chr [1:90] "cadillac" "cadillac" "cadillac" "cadillac" ...  
## $ model : chr [1:90] "escalade" "escalade esv" "escalade esv" "escalade" ...  
## $ condition : chr [1:90] "excellent" "good" "good" "good" ...  
## $ cylinders : num [1:90] 8 8 8 8 8 6 8 8 8 6 ...  
## $ fuel : chr [1:90] "gas" "gas" "gas" "gas" ...  
## $ odometer : num [1:90] 240000 112601 205023 108178 78000 ...  
## $ paint.color: chr [1:90] "white" "white" "red" "white" ...

#sample test if the year is 2001  
s90$year == 2001

## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [85] FALSE FALSE FALSE FALSE FALSE FALSE

#sample test if the paint.color is black  
  
s90$paint.color == "black"

## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [85] FALSE FALSE FALSE FALSE FALSE FALSE

#similarlly we can perform others also  
  
#⦁ Use R to conduct a multiple linear regression on your random sample with PRICE as the dependent variable and ODOMETER, YEAR, CONDITION, PAINT COLOR and CYLINDERS as the independent variables. Make appropriate decisions as to whether any of the independent variables should be treated as factor variables.   
  
price.out = lm(price~odometer+year+condition+paint.color+cylinders,data=s90)  
  
summary(price.out)

##   
## Call:  
## lm(formula = price ~ odometer + year + condition + paint.color +   
## cylinders, data = s90)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7768.6 -1964.3 -64.1 1644.3 9114.0   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.775e+06 5.127e+05 -5.413 6.47e-07 \*\*\*  
## odometer -1.508e-02 9.820e-03 -1.536 0.129   
## year 1.380e+03 2.547e+02 5.419 6.32e-07 \*\*\*  
## conditiongood 2.540e+02 8.769e+02 0.290 0.773   
## paint.colorbrown 7.463e+02 2.450e+03 0.305 0.762   
## paint.colorgrey -6.490e+02 1.872e+03 -0.347 0.730   
## paint.colorpurple 6.672e+02 3.055e+03 0.218 0.828   
## paint.colorred 1.428e+03 1.792e+03 0.797 0.428   
## paint.colorsilver -3.520e+02 1.664e+03 -0.212 0.833   
## paint.colorwhite 1.263e+03 1.693e+03 0.746 0.458   
## cylinders 2.108e+03 4.813e+02 4.379 3.62e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3737 on 79 degrees of freedom  
## Multiple R-squared: 0.4183, Adjusted R-squared: 0.3447   
## F-statistic: 5.682 on 10 and 79 DF, p-value: 2.313e-06

#leaving out the paint.color because we are failing to reject the null hypothesis and the p-value of these is very greater than 0.05  
  
price.out = lm(price~odometer+year+condition+cylinders,data=s90)  
summary(price.out)

##   
## Call:  
## lm(formula = price ~ odometer + year + condition + cylinders,   
## data = s90)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7071.9 -2226.1 -246.6 1814.0 9922.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.809e+06 4.693e+05 -5.986 4.97e-08 \*\*\*  
## odometer -1.511e-02 9.525e-03 -1.586 0.116   
## year 1.396e+03 2.330e+02 5.992 4.84e-08 \*\*\*  
## conditiongood 3.007e+02 8.360e+02 0.360 0.720   
## cylinders 2.258e+03 4.207e+02 5.368 6.79e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3699 on 85 degrees of freedom  
## Multiple R-squared: 0.3867, Adjusted R-squared: 0.3578   
## F-statistic: 13.4 on 4 and 85 DF, p-value: 1.651e-08

#leaving out the condition with the same above reason  
  
price.out = lm(price~odometer+year+cylinders,data=s90)  
summary(price.out)

##   
## Call:  
## lm(formula = price ~ odometer + year + cylinders, data = s90)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7128.3 -2101.7 -194.3 1934.8 9866.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.788e+06 4.632e+05 -6.019 4.18e-08 \*\*\*  
## odometer -1.497e-02 9.469e-03 -1.581 0.118   
## year 1.386e+03 2.300e+02 6.025 4.07e-08 \*\*\*  
## cylinders 2.246e+03 4.172e+02 5.384 6.23e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3681 on 86 degrees of freedom  
## Multiple R-squared: 0.3858, Adjusted R-squared: 0.3643   
## F-statistic: 18 on 3 and 86 DF, p-value: 3.736e-09

#leaving out the odometer with the same above reason  
  
price.out = lm(price~year+cylinders,data=s90)  
  
summary(price.out)

##   
## Call:  
## lm(formula = price ~ year + cylinders, data = s90)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6915.4 -2582.8 -657.7 2219.2 10389.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2924154.2 458945.5 -6.371 8.61e-09 \*\*\*  
## year 1453.2 228.0 6.374 8.51e-09 \*\*\*  
## cylinders 2175.5 418.3 5.200 1.30e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3712 on 87 degrees of freedom  
## Multiple R-squared: 0.3679, Adjusted R-squared: 0.3534   
## F-statistic: 25.32 on 2 and 87 DF, p-value: 2.157e-09

#now the model has good beta co efficients and the p-values are close to zero, so now lets move to independent variables depending on other independent variable  
  
price.out = lm(price~year+cylinders+I(year\*cylinders),data=s90)  
summary(price.out)

##   
## Call:  
## lm(formula = price ~ year + cylinders + I(year \* cylinders),   
## data = s90)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7008.1 -2667.5 -617.2 2281.6 10296.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -2.466e+06 3.308e+06 -0.746 0.458  
## year 1.225e+03 1.647e+03 0.744 0.459  
## cylinders -6.245e+04 4.622e+05 -0.135 0.893  
## I(year \* cylinders) 3.218e+01 2.301e+02 0.140 0.889  
##   
## Residual standard error: 3733 on 86 degrees of freedom  
## Multiple R-squared: 0.3681, Adjusted R-squared: 0.346   
## F-statistic: 16.7 on 3 and 86 DF, p-value: 1.24e-08

#the values very greater than 0.05 so not dependency  
  
price.out = lm(price~year+cylinders+I(year^2),data=s90)  
summary(price.out)

##   
## Call:  
## lm(formula = price ~ year + cylinders + I(year^2), data = s90)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6518.0 -2450.4 -700.4 2642.8 10787.0   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.795e+08 6.384e+08 -0.908 0.367   
## year 5.756e+05 6.357e+05 0.905 0.368   
## cylinders 2.193e+03 4.192e+02 5.232 1.17e-06 \*\*\*  
## I(year^2) -1.429e+02 1.583e+02 -0.903 0.369   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3716 on 86 degrees of freedom  
## Multiple R-squared: 0.3738, Adjusted R-squared: 0.352   
## F-statistic: 17.12 on 3 and 86 DF, p-value: 8.405e-09

#can reject this model with the same reason  
  
price.out = lm(price~year+cylinders+I(cylinders^2),data=s90)  
summary(price.out)

##   
## Call:  
## lm(formula = price ~ year + cylinders + I(cylinders^2), data = s90)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6915.4 -2582.8 -657.7 2219.2 10389.6   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2924154.2 458945.5 -6.371 8.61e-09 \*\*\*  
## year 1453.2 228.0 6.374 8.51e-09 \*\*\*  
## cylinders 2175.5 418.3 5.200 1.30e-06 \*\*\*  
## I(cylinders^2) NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3712 on 87 degrees of freedom  
## Multiple R-squared: 0.3679, Adjusted R-squared: 0.3534   
## F-statistic: 25.32 on 2 and 87 DF, p-value: 2.157e-09

#no effect   
  
  
  
price.out = lm(price~year+cylinders+I(year\*cylinders),data=s90)  
summary(price.out)

##   
## Call:  
## lm(formula = price ~ year + cylinders + I(year \* cylinders),   
## data = s90)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7008.1 -2667.5 -617.2 2281.6 10296.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -2.466e+06 3.308e+06 -0.746 0.458  
## year 1.225e+03 1.647e+03 0.744 0.459  
## cylinders -6.245e+04 4.622e+05 -0.135 0.893  
## I(year \* cylinders) 3.218e+01 2.301e+02 0.140 0.889  
##   
## Residual standard error: 3733 on 86 degrees of freedom  
## Multiple R-squared: 0.3681, Adjusted R-squared: 0.346   
## F-statistic: 16.7 on 3 and 86 DF, p-value: 1.24e-08

price.out = lm(price~year+cylinders,data=s90)  
summary(price.out)

##   
## Call:  
## lm(formula = price ~ year + cylinders, data = s90)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6915.4 -2582.8 -657.7 2219.2 10389.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2924154.2 458945.5 -6.371 8.61e-09 \*\*\*  
## year 1453.2 228.0 6.374 8.51e-09 \*\*\*  
## cylinders 2175.5 418.3 5.200 1.30e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3712 on 87 degrees of freedom  
## Multiple R-squared: 0.3679, Adjusted R-squared: 0.3534   
## F-statistic: 25.32 on 2 and 87 DF, p-value: 2.157e-09

#condition was already considered as factor but was ignored later because the fit has higher p-values which is not good now let us consider the cylinders as factor and try  
  
s90f = s90  
s90f$cylinders = as.factor(s90f$cylinders)  
attach(s90f)

## The following objects are masked from M5 (pos = 3):  
##   
## condition, cylinders, fuel, make, model, odometer, paint.color,  
## price, region, year

## The following objects are masked from M5 (pos = 4):  
##   
## condition, cylinders, fuel, make, model, odometer, paint.color,  
## price, region, year

price1.out = lm(price~odometer+year+condition+paint.color+cylinders,data=s90f)  
summary(price1.out)

##   
## Call:  
## lm(formula = price ~ odometer + year + condition + paint.color +   
## cylinders, data = s90f)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7768.6 -1964.3 -64.1 1644.3 9114.0   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.763e+06 5.116e+05 -5.400 6.82e-07 \*\*\*  
## odometer -1.508e-02 9.820e-03 -1.536 0.129   
## year 1.380e+03 2.547e+02 5.419 6.32e-07 \*\*\*  
## conditiongood 2.540e+02 8.769e+02 0.290 0.773   
## paint.colorbrown 7.463e+02 2.450e+03 0.305 0.762   
## paint.colorgrey -6.490e+02 1.872e+03 -0.347 0.730   
## paint.colorpurple 6.672e+02 3.055e+03 0.218 0.828   
## paint.colorred 1.428e+03 1.792e+03 0.797 0.428   
## paint.colorsilver -3.520e+02 1.664e+03 -0.212 0.833   
## paint.colorwhite 1.263e+03 1.693e+03 0.746 0.458   
## cylinders8 4.215e+03 9.626e+02 4.379 3.62e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3737 on 79 degrees of freedom  
## Multiple R-squared: 0.4183, Adjusted R-squared: 0.3447   
## F-statistic: 5.682 on 10 and 79 DF, p-value: 2.313e-06

#repeating similarly  
price1.out = lm(price~odometer+year+condition+cylinders,data=s90f)  
summary(price1.out)

##   
## Call:  
## lm(formula = price ~ odometer + year + condition + cylinders,   
## data = s90f)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7071.9 -2226.1 -246.6 1814.0 9922.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.796e+06 4.685e+05 -5.968 5.38e-08 \*\*\*  
## odometer -1.511e-02 9.525e-03 -1.586 0.116   
## year 1.396e+03 2.330e+02 5.992 4.84e-08 \*\*\*  
## conditiongood 3.007e+02 8.360e+02 0.360 0.720   
## cylinders8 4.517e+03 8.414e+02 5.368 6.79e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3699 on 85 degrees of freedom  
## Multiple R-squared: 0.3867, Adjusted R-squared: 0.3578   
## F-statistic: 13.4 on 4 and 85 DF, p-value: 1.651e-08

price1.out = lm(price~odometer+year+cylinders,data=s90f)  
summary(price1.out)

##   
## Call:  
## lm(formula = price ~ odometer + year + cylinders, data = s90f)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7128.3 -2101.7 -194.3 1934.8 9866.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.774e+06 4.623e+05 -6.000 4.53e-08 \*\*\*  
## odometer -1.497e-02 9.469e-03 -1.581 0.118   
## year 1.386e+03 2.300e+02 6.025 4.07e-08 \*\*\*  
## cylinders8 4.492e+03 8.343e+02 5.384 6.23e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3681 on 86 degrees of freedom  
## Multiple R-squared: 0.3858, Adjusted R-squared: 0.3643   
## F-statistic: 18 on 3 and 86 DF, p-value: 3.736e-09

price1.out = lm(price~year+cylinders,data=s90f)  
summary(price1.out)

##   
## Call:  
## lm(formula = price ~ year + cylinders, data = s90f)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6915.4 -2582.8 -657.7 2219.2 10389.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2911101.5 458056.1 -6.355 9.25e-09 \*\*\*  
## year 1453.2 228.0 6.374 8.51e-09 \*\*\*  
## cylinders8 4350.9 836.6 5.200 1.30e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3712 on 87 degrees of freedom  
## Multiple R-squared: 0.3679, Adjusted R-squared: 0.3534   
## F-statistic: 25.32 on 2 and 87 DF, p-value: 2.157e-09

price.out = lm(price~year+cylinders,data=s90)  
summary(price.out)

##   
## Call:  
## lm(formula = price ~ year + cylinders, data = s90)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6915.4 -2582.8 -657.7 2219.2 10389.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2924154.2 458945.5 -6.371 8.61e-09 \*\*\*  
## year 1453.2 228.0 6.374 8.51e-09 \*\*\*  
## cylinders 2175.5 418.3 5.200 1.30e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3712 on 87 degrees of freedom  
## Multiple R-squared: 0.3679, Adjusted R-squared: 0.3534   
## F-statistic: 25.32 on 2 and 87 DF, p-value: 2.157e-09

#here we reached out to the same equation with different beta co-efficients now we need consider the beta with low value if our buying the car because we get for low value and for selling the car we need to take high beta.  
  
# Let us consider the factor as our primary scenario.  
#that is price1.out is our model  
  
#⦁ Report the beta coefficients and associated p values and beta coefficient confidence intervals from your model. Provide appropriate interpretations of your beta coefficients.  
  
confint(price1.out)

## 2.5 % 97.5 %  
## (Intercept) -3821537.594 -2000665.449  
## year 1000.041 1906.335  
## cylinders8 2688.000 6013.814

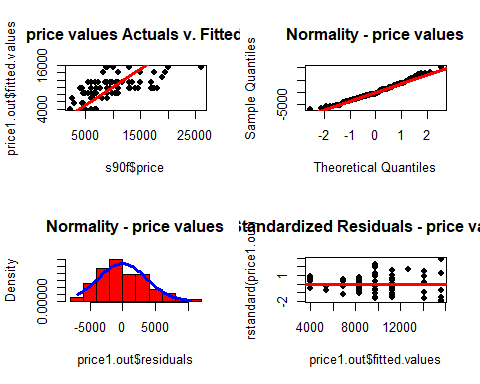
summary(price1.out)

##   
## Call:  
## lm(formula = price ~ year + cylinders, data = s90f)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6915.4 -2582.8 -657.7 2219.2 10389.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2911101.5 458056.1 -6.355 9.25e-09 \*\*\*  
## year 1453.2 228.0 6.374 8.51e-09 \*\*\*  
## cylinders8 4350.9 836.6 5.200 1.30e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3712 on 87 degrees of freedom  
## Multiple R-squared: 0.3679, Adjusted R-squared: 0.3534   
## F-statistic: 25.32 on 2 and 87 DF, p-value: 2.157e-09

#y=-2911101.5+1453.2(b1)+4350.9(b2)  
  
# this is the equation  
  
#for every increase in 1 year the price of the car increases by 1453.2 and this value is not perfect and the increase lies in confidence interval of (1000.041 1906.335)  
#for cylinders increase from 6 to 8 the price of the car increases by 4350.9 and this value is not perfect and the increase lies in confidence interval of (2688.000 6013.814)  
  
#⦁ Conduct appropriate analyses and give interpretations to determine if your model is a good fit to the data in your primary data set.   
  
summary(price1.out)

##   
## Call:  
## lm(formula = price ~ year + cylinders, data = s90f)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6915.4 -2582.8 -657.7 2219.2 10389.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2911101.5 458056.1 -6.355 9.25e-09 \*\*\*  
## year 1453.2 228.0 6.374 8.51e-09 \*\*\*  
## cylinders8 4350.9 836.6 5.200 1.30e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3712 on 87 degrees of freedom  
## Multiple R-squared: 0.3679, Adjusted R-squared: 0.3534   
## F-statistic: 25.32 on 2 and 87 DF, p-value: 2.157e-09

#all the p-values for our beta coefficients are very low which is a good assumption but the factor cylinders is effecting more than the other all variables in the data . Also the confidence intervals of the slopes have very high range which is not a good thing.  
  
#⦁ Assess your model’s conformance with the LINE assumptions of regression.  
  
par(mfrow=c(2,2))  
#Linearity  
plot(s90f$price,price1.out$fitted.values,pch=19,main="price values Actuals v. Fitted")  
abline(0,1,lwd=3,col="red")  
  
# seems to be a pattern with equal weights on both sides  
#linearity is fine  
  
#Normality  
qqnorm(price1.out$residuals,pch=19,  
 main="Normality - price values")  
qqline(price1.out$residuals,lwd=3,col="red")  
hist(price1.out$residuals,col="red",probability = TRUE,  
 main="Normality - price values")  
curve(dnorm(x,0,sd(price1.out$residuals)),  
 from=min(price1.out$residuals),  
 to=max(price1.out$residuals),  
 lwd=3,col="blue",add=TRUE)  
  
# most of the points lie on the QQline and the hist graph also shows that it is normal distribution  
#normality is fine  
  
#Equality of Variances  
plot(price1.out$fitted.values,rstandard(price1.out),pch=19,  
 main="Standardized Residuals - price values")  
abline(0,0,col="red",lwd=3)



#there is difference here with low price cars having low sd and then increases and goes on  
#equality of variances is a problem because few points have reached 2 sd's line and others did not.  
  
#⦁ Throckmorton P. Gildersleeve of Summerfield, Vermont would like to sell his red 2011 Cadillac DTS pictured above. He says the vehicle is in "excellent" condition and has 215,354 miles on the odometer. Mr. Gildersleeve has not shared details of his Cadillac’s engine with you because he thinks you know that all 2011 DTS cars had the same famous engine. If you know nothing about Cadillac engines the point is easily researched online. Use R and your model to determine what price he should ask for the car. Do you believe your pricing advice to the Great Gildersleeve is accurate and usable? Give reasoning for your conclusion.  
  
newdata =data.frame(year=2011,cylinders ='8',odometer=215354,condition = 'execellent')  
  
predict(price1.out,newdata,interval = "predict")

## fit lwr upr  
## 1 15610.44 8018.685 23202.2

#my model has the selling price of 15610.44 but it can range from 8018.685 to 23202.2 which is very high variance and also the R-squared value is below 50% which cant explain the dependency