

# 603 Project - Car selling price predictor

Team-1

04/12/2021

## Introduction

The Indian used car market is expected to reach a size of over 70 lakh vehicles by 2025-26, up from 38 lakh in 2020-21, growing at around 12-14 per cent over the next few years, according to a study. The used car industry is expected to clock a growth rate of 15 per cent in FY22 as the COVID-19 pandemic, digitization, changing demographics and aspirations, first-time buyers, and availability of financing options acting as growth drivers, as per the OLX Autos-CRISIL Study 2021.

### Research question

Investigating the factors that have a statistical significance on the price of used car.

### Expected finding

To develop a best fit model that predict the price of car given the required predictor variables.

### Topic

Being a student, it is difficult to afford a new car and most of the time they choose to buy a second-hand car. Also, students who have very little knowledge about cars find it difficult to figure out if the price for the car is fair based on its specifications.

### Data

The data is being collected from Kaggle and the data is scraped from cardekho.com. It contains 8100 rows and 13 columns.

## Methodology

### Descriptive Statistics

**Response Variable:** selling\_price – This variable has the selling price of the used cars in CAD.

### Independent Variables:

- **Quantitative variables:**

1. km\_driven – Total distance the car has travelled. The units are Kilometres.

2. Mileage – The distance that a car travels per 1 litre of fuel. The units are in km/ltr
3. Engine – The space/volume available inside the tank to accommodate air-fuel mix for burning. The units are in CC (cubic capacity)
4. Max\_power – The power available at the shaft of an engine. The unit is bhp (brake horsepower)
5. Age – This variable shows how old the car is in number of years.

- **Qualitative variables:**

1. Fuel - The type of fuel the car runs on. Types: Petrol, diesel, CNG, LPG
2. Seller-type – Whether it's being sold by a dealer or an individual.
3. Transmission – There are 2 types: Manual or Automatic
4. Owner – By how many people the car has been owned. Types: First, Second, Third, Fourth & Above or Test drive car
5. Seats – Number of seats present in the vehicle

## **Modeling**

Part 1 - Data cleaning and wrangling

1. Import the CSV file in R
2. Removed the columns ‘name’ and torque’
3. Removed the rows that contained ‘km/kg’ in the ‘mileage’ column
4. All the units present in the ‘mileage’, ‘engine’, and ‘max\_power’ is removed
5. The age of the car is calculated by taking the difference between 2020 and the ‘year’ of the car. This is stored in the column ‘age’ 6.The columns ‘mileage’, ‘engine’, and ‘max\_power’ are changed to numeric values
6. All ‘NA’ values are removed
7. The ‘selling\_price’ is in Rupees. This is changed to CAD

Part 2 - Modeling

1. Plotting relationship between each individual variable and response variable.
2. Checking all pairwise combinations of predictors in scatter plots and using the VIF function(Test the multi-collinearity assumption)
3. Implement the full model
4. Use step wise regression (with backward elimination) to find the “best” set of predictors of selling\_price
5. Use all-possible-regressions-selection to find the “best” predictors of selling\_price (Cp, AIC, Adjusted R2)

6. Test the hypothesis for the full model.
7. Conducting a partial F-test, to test if number of seats should be kept in the model.
8. Build an interaction model to fit the multiple regression model from the model.
9. Conduct a partial t-test to check for significance of interaction term.
10. Create a model fit which contains higher degrees.
11. Compare the R  $\hat{R}$  value and RMSE value of the additive model, interaction model and higher model.
12. Test the best model for model assumptions such as equal variance, linearity, normality.
13. Check for outliers using the cook's distance.
14. Model transformation
15. Finalize the model. Predict the selling\_price.

## Part 1 - Data cleaning and wrangling

### 1. Import the CSV file in R

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr    1.0.7
## v tidyr   1.1.3     v stringr  1.4.0
## v readr   2.0.1     vforcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()

car = read.csv("/Users/ohamugochukwu/Desktop/Car (1).csv")
```

### 2. Removed the columns ‘name’ and torque’

```
car1 = subset(car, select = -c(name, torque))
```

### 3. Removed the rows that contained ‘km/kg’ in the ‘mileage’ column

```
car1 = car1[!grepl("km/kg", car1$mileage),]
```

4. All the units present in the ‘mileage’, ‘engine’, and ‘max\_power’ is removed
5. The age of the car is calculated by taking the difference between 2020 and the ‘year’ of the car. This is stored in the column ‘age’

```

library(stringr)
car1$mileage = str_split_fixed(car1$mileage, " ", 2)[ ,1]
car1$engine = str_split_fixed(car1$engine, " ", 2)[ ,1]
car1$max_power = str_split_fixed(car1$max_power, " ", 2)[ ,1] # remove the
  ↳ units from the column
car1$year = 2020 - car1$year

```

```

car1 = car1 %>% rename(age= year) #rename the year to age. To show age of
  ↳ the car

```

6. The columns ‘mileage’, ‘engine’, and ‘max\_power’ are changed to numeric values

7. All ‘NA’ values are removed

8. The ‘selling\_price’ is in Rupees. This is changed to CAD

```

car1$mileage = as.numeric(car1$mileage)
car1$engine = as.numeric(car1$engine)
car1$max_power = as.numeric(car1$max_power) # convert the values to numeric
car1 = na.omit(car1) #remove na values
car1$selling_price = car1$selling_price/60 #convert price to CAD dollars
head(car1)

```

```

##   age selling_price km_driven fuel seller_type transmission          owner
## 1   6     7500.000    145500 Diesel Individual      Manual First Owner
## 2   6     6166.667    120000 Diesel Individual      Manual Second Owner
## 3  14     2633.333    140000 Petrol Individual      Manual Third Owner
## 4  10     3750.000    127000 Diesel Individual      Manual First Owner
## 5  13     2166.667    120000 Petrol Individual      Manual First Owner
## 6   3     7333.333     45000 Petrol Individual      Manual First Owner
##   mileage engine max_power seats
## 1   23.40    1248     74.00     5
## 2   21.14    1498    103.52     5
## 3   17.70    1497     78.00     5
## 4   23.00    1396     90.00     5
## 5   16.10    1298     88.20     5
## 6   20.14    1197     81.86     5

```

```

Car1 = car1

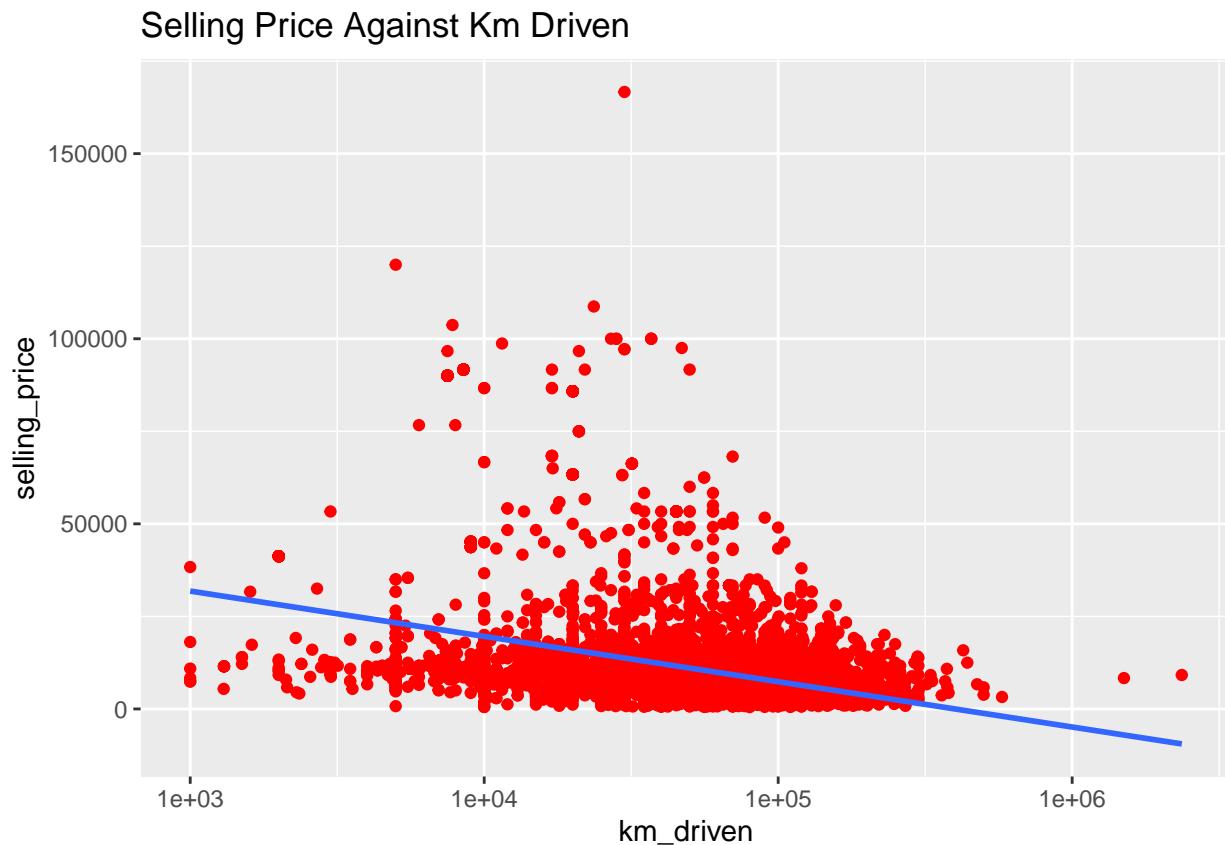
```

## Part 2 - Modeling

1. Plotting relationship between each individual variable and response variable.

```
ggplot(data=Car1,mapping=
  aes(x=km_driven,y=selling_price))+geom_point(color='red')+  
  geom_smooth(method = "lm", se = FALSE)+ggtitle("Selling Price Against Km  
  Driven")+
  scale_x_log10()
```

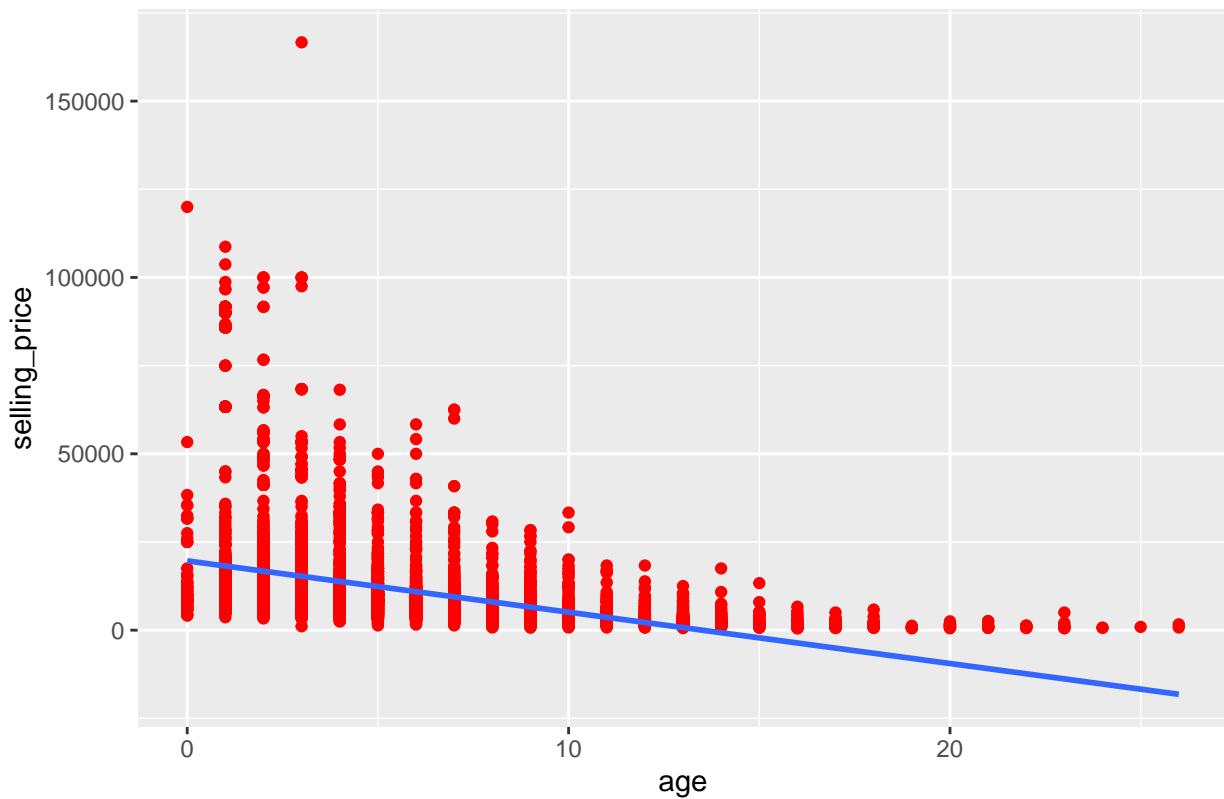
```
## `geom_smooth()` using formula 'y ~ x'
```



The scatter plot above, shows that there is an inverse linear realtionship between the the selling price of a car and the distance in kilometers the car has covered.

```
ggplot(data=Car1,mapping=
  aes(x=age,y=selling_price))+geom_point(color='red')+  
  geom_smooth(method = "lm", se = FALSE)+ggtitle("Selling Price Against Age")  
  
## `geom_smooth()` using formula 'y ~ x'
```

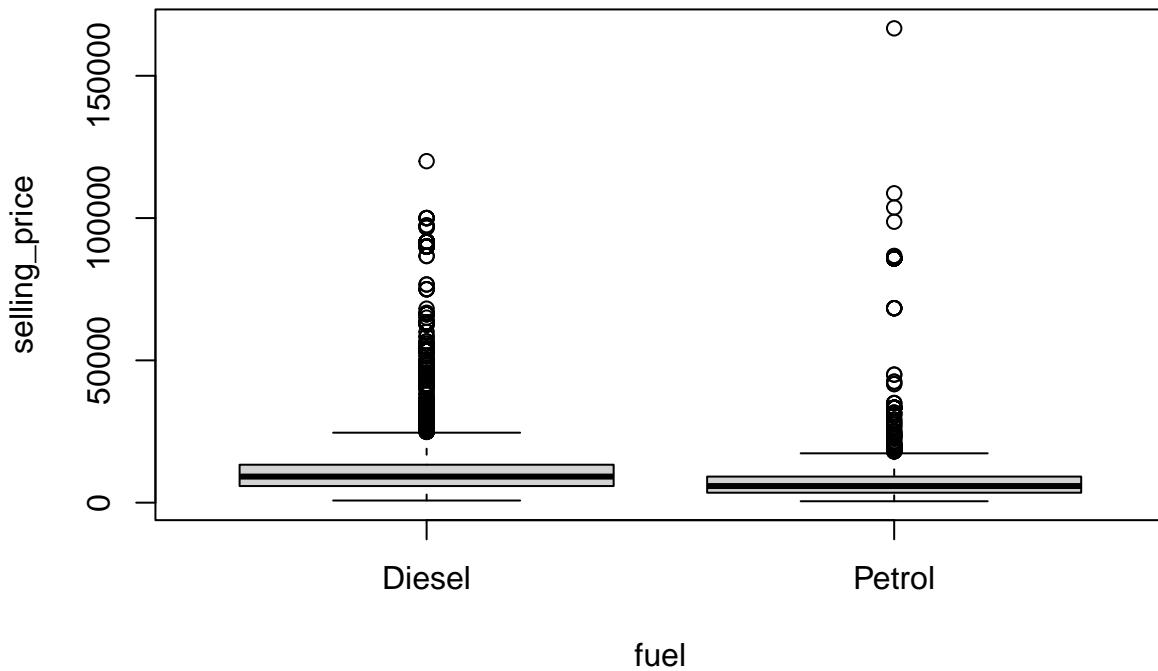
## Selling Price Against Age



The scatter plot above shows that there is an inverse linear relationship between the selling price of a car and the age of the car.

```
#ggplot(data=Car1,mapping= aes(x=fuel,y=selling_price))+ geom_boxplot()
boxplot(selling_price~fuel,data=Car1,
        main = "Boxplot of Selling Price Against Fuel Type")
```

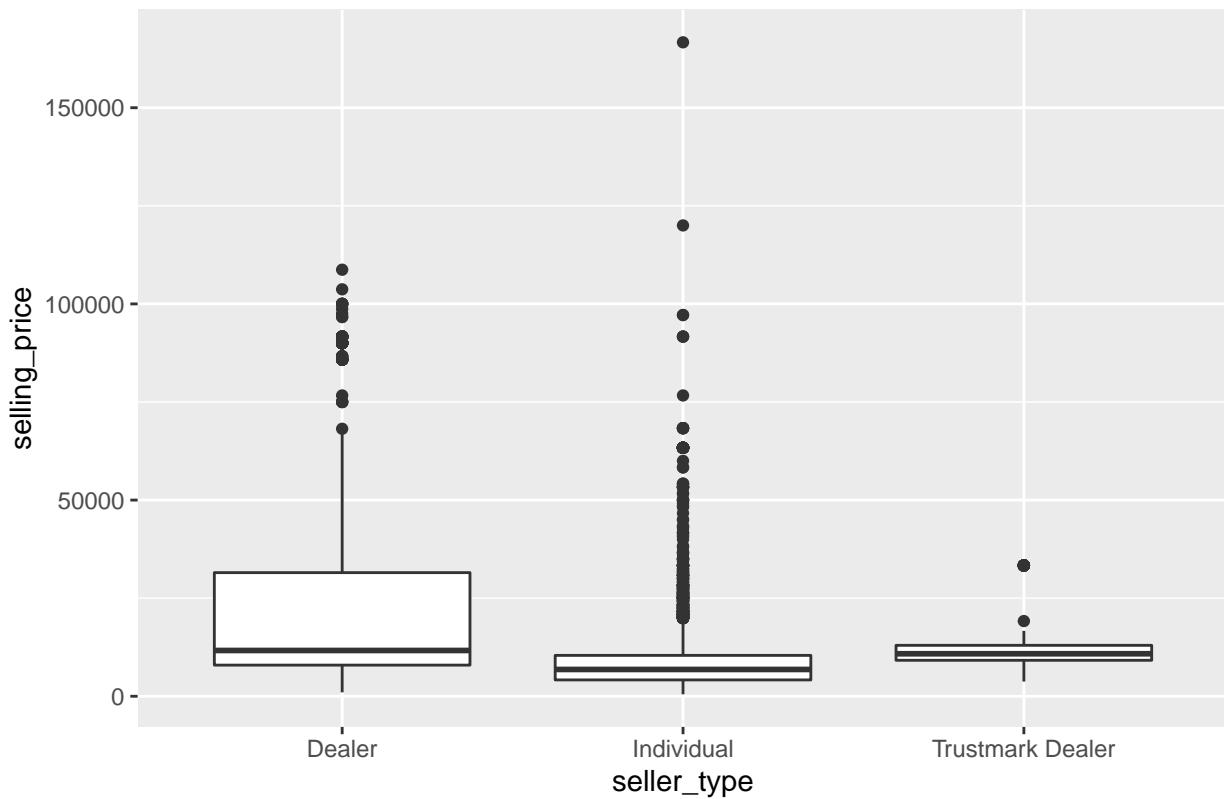
## Boxplot of Selling Price Against Fuel Type



From the boxplot diagram above, the median of the cars that run on diesel, is slightly higher than cars that run on petrol. Although the diesel data has more outliers.

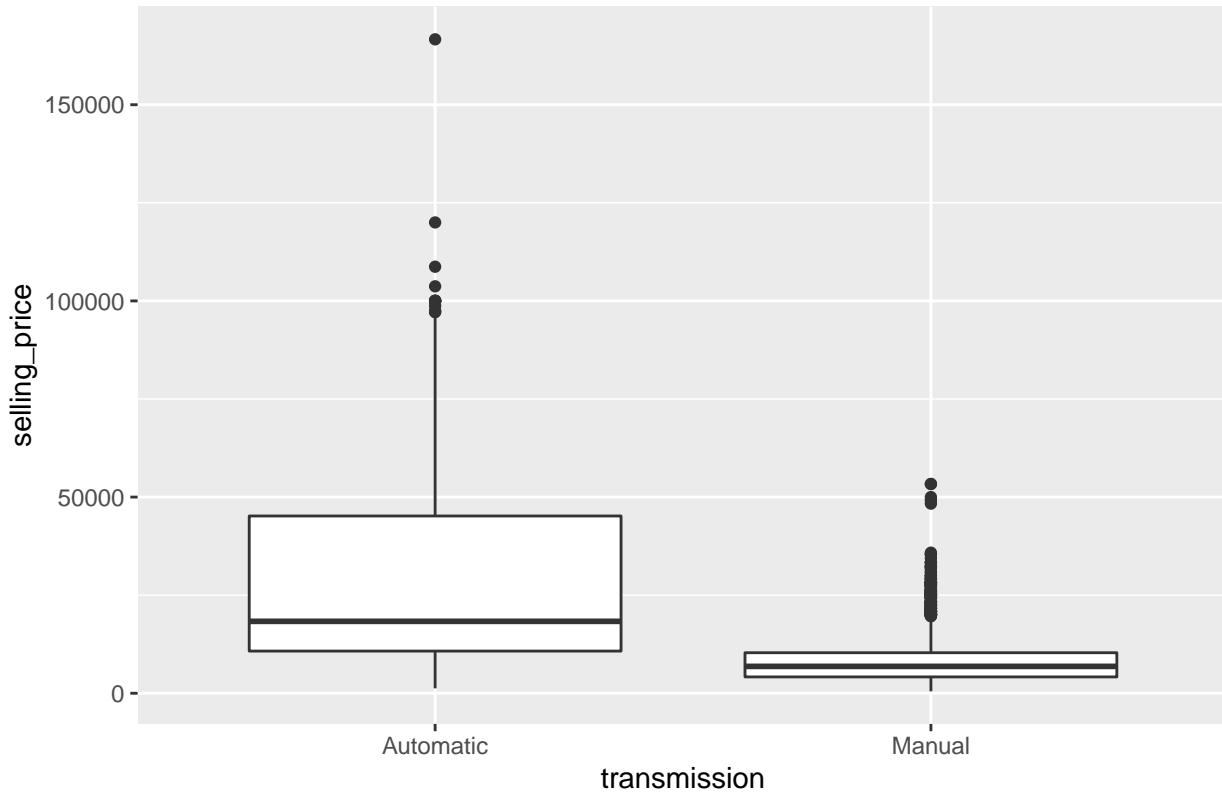
```
ggplot(data=Car1,mapping= aes(x=seller_type,y=selling_price))+  
  geom_boxplot() +  
  ggtitle("Boxplot of Selling Price Against Seller Type")
```

Boxplot of Selling Price Against Seller Type



```
ggplot(data=Car1,mapping= aes(x=transmission,y=selling_price))+
  geom_boxplot()+
  ggtitle("Boxplot of Selling Price Against Transmission")
```

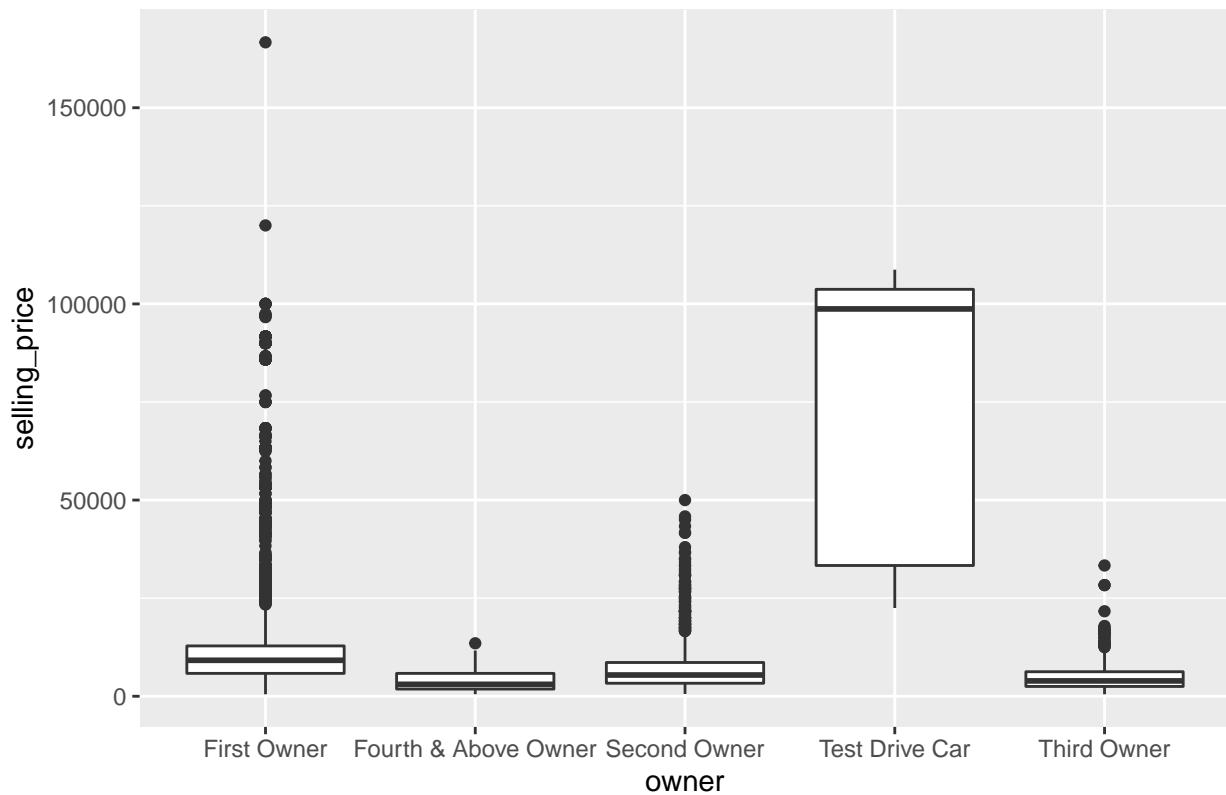
Boxplot of Selling Price Against Transmission



The boxplot diagram above indicates that the median of cars with automatic transmission, is higher than that with manual transmission.

```
ggplot(data=Car1,mapping= aes(x=owner,y=selling_price))+ geom_boxplot()+
  ggtitle("Boxplot of Selling Price Against Owner")
```

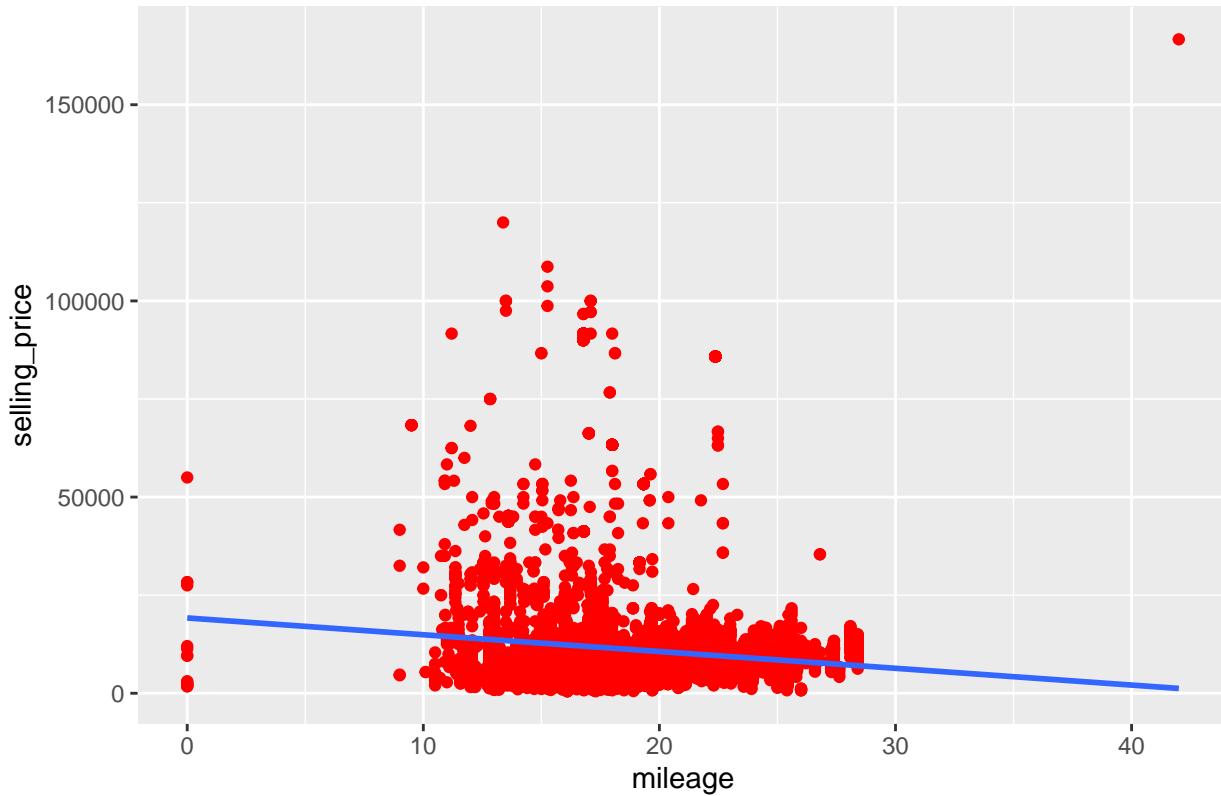
Boxplot of Selling Price Against Owner



```
ggplot(data=Car1,mapping=
  ↪ aes(x=mileage,y=selling_price))+geom_point(color='red')+  
  geom_smooth(method = "lm", se = FALSE)+ggtitle("Selling Price Against  
  ↪ Mileage")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

## Selling Price Against Mileage

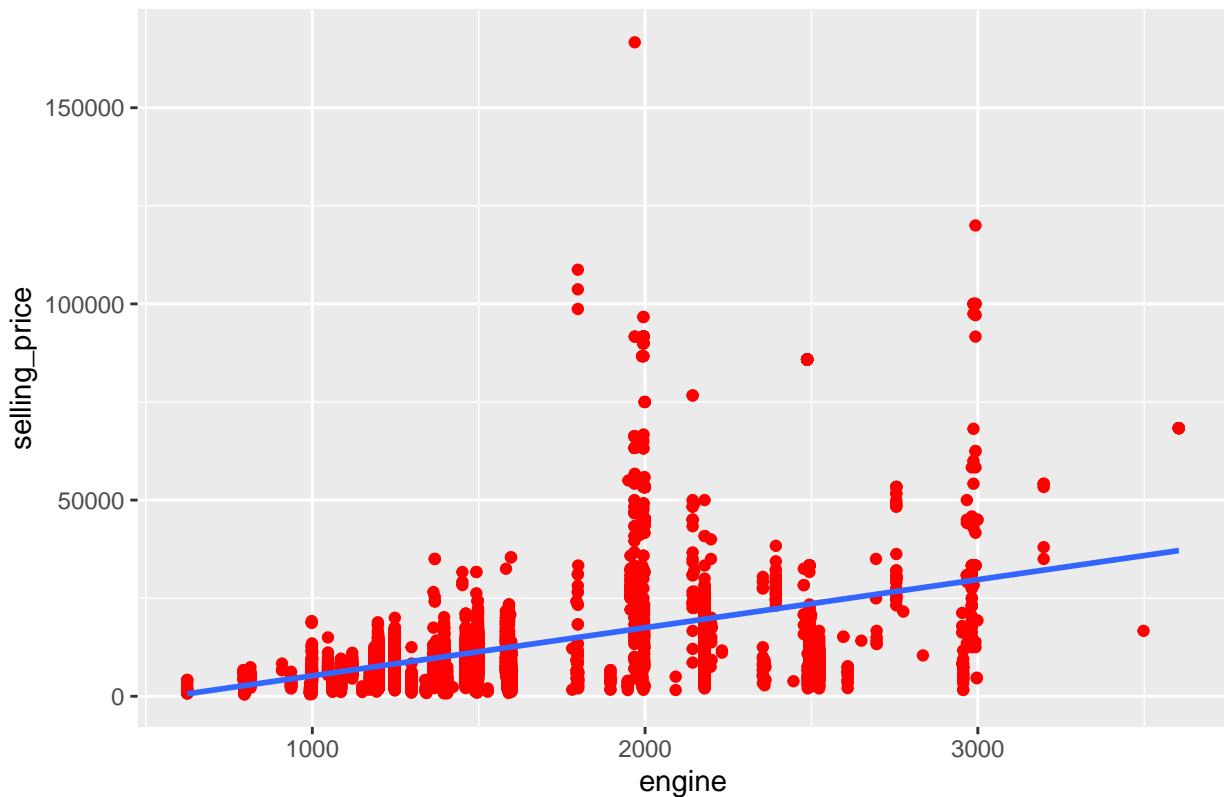


From the scattered plot above, the selling price of a used car has an inverse relationship with the mileage of the car. That is, as the mileage increases, the selling price of the car decreases.

```
ggplot(data=Car1,mapping=
  ↪   aes(x=engine,y=selling_price))+geom_point(color='red')+
  geom_smooth(method = "lm", se = FALSE)+ggtitle("Selling Price Against
  ↪   Engine")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

## Selling Price Against Engine

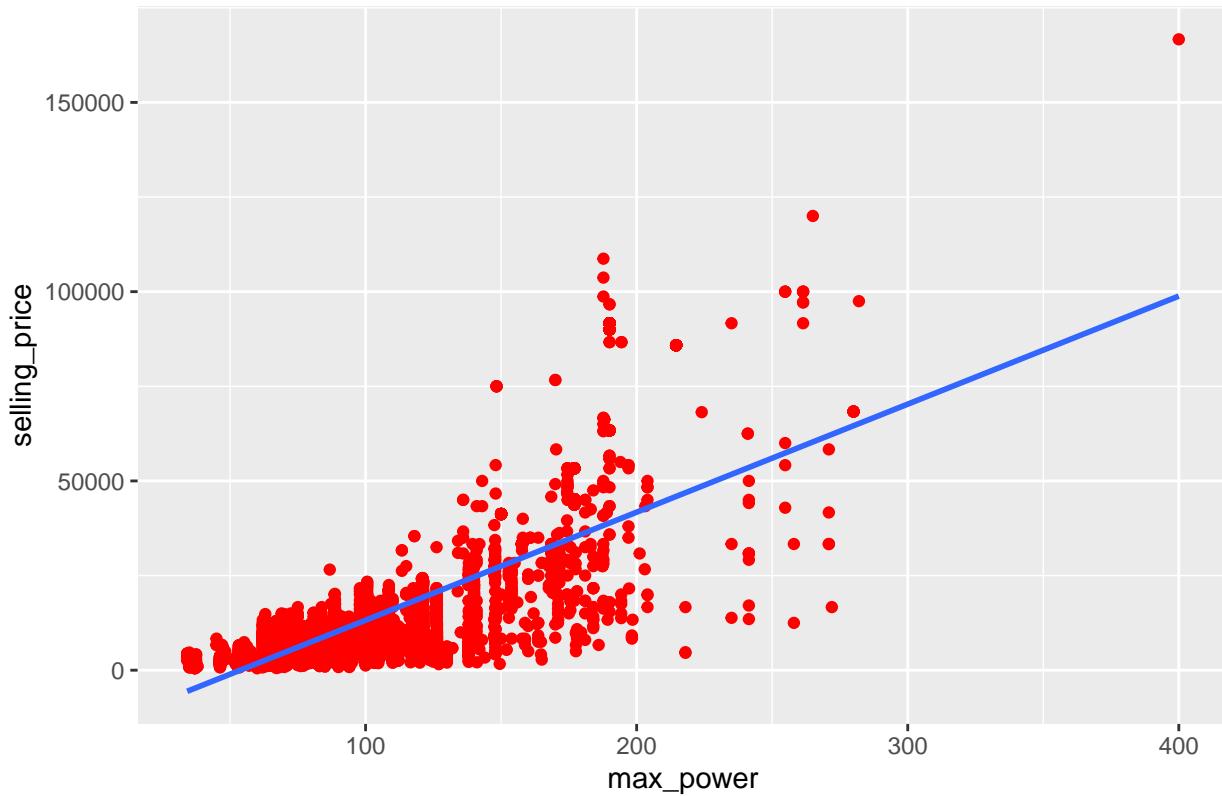


From the scattered plot above, the selling price of a used car has a positive relationship with the engine of the car. That is, as the engine capacity increases, the selling price of the car increases.

```
ggplot(data=Car1,mapping=
  ↪ aes(x=max_power,y=selling_price))+geom_point(color='red')+
  geom_smooth(method = "lm", se = FALSE)+ggtitle("Selling Price Against
  ↪ Maximum Power")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

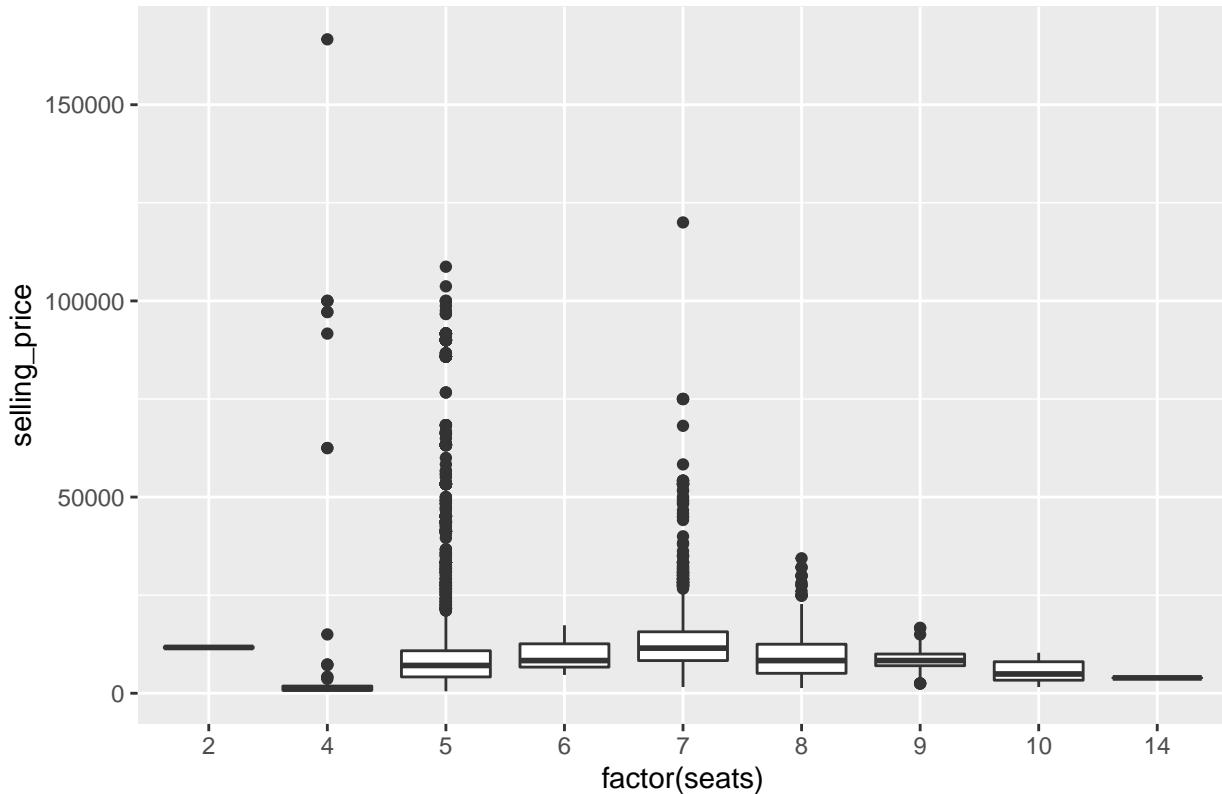
### Selling Price Against Maximum Power



The scattered plot diagram above shows the positive linear relationship between the selling price of a used car and the maximum power produced by the car.

```
ggplot(data=Car1,mapping= aes(x=factor(seats),y=selling_price))+  
  geom_boxplot() +  
  ggtitle("Boxplot of Selling Price Against no of Seats")
```

Boxplot of Selling Price Against no of Seats



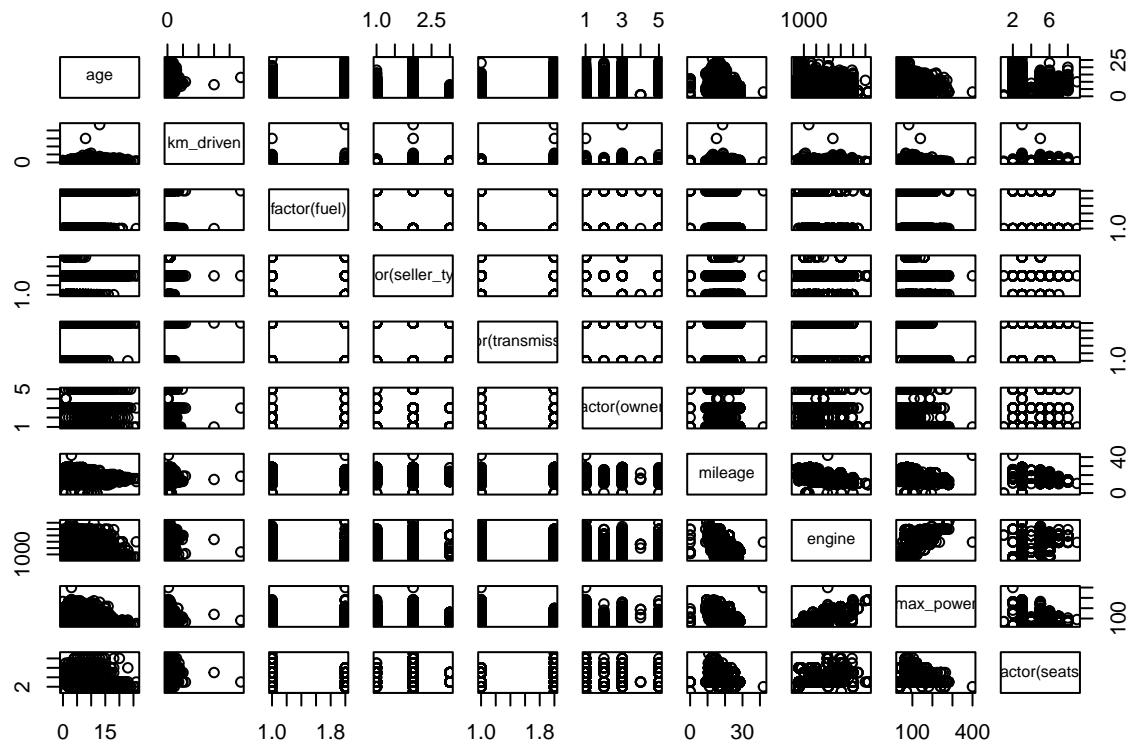
2. Checking all pairwise combinations of predictors in scatter plots and using the VIF function (Test the multi-collinearity assumption)

```
#Test the model for multicollinearity
firstordermodel = lm(selling_price ~ age + km_driven + factor(fuel) +
  ↪ factor(seller_type)
  + factor(transmission)+ factor(owner)+ mileage +
  ↪ engine + max_power+
  factor(seats),data = Car1)
```

```
#pairwise combinations
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
pairs(~age + km_driven + factor(fuel) + factor(seller_type)
  + factor(transmission)+ factor(owner)+ mileage +
  ↪ engine + max_power+
  factor(seats),data = Car1,)
```



```
#ggpairs(firstordermodel, lower =list(continuous ="smooth_loess",
#combo ="facethist", discrete ="facetbar", na ="na"))
```

```
# VIF function
library(mctest)
imcdiag(firstordermodel, method = "VIF")
```

```
##
## Call:
## imcdiag(mod = firstordermodel, method = "VIF")
##
##
## VIF Multicollinearity Diagnostics
##
##                                     VIF detection
## age                           2.2017      0
## km_driven                     1.4560      0
## factor(fuel)Petrol             2.1693      0
## factor(seller_type)Individual  1.4492      0
## factor(seller_type)Trustmark Dealer 1.2483      0
```

```

## factor(transmission)Manual          1.7056      0
## factor(owner)Fourth & Above Owner 1.1017      0
## factor(owner)Second Owner         1.2823      0
## factor(owner)Test Drive Car       1.0067      0
## factor(owner)Third Owner          1.2019      0
## mileage                           2.9450      0
## engine                            5.5399      0
## max_power                          3.3690      0
## factor(seats)4                   66.5178     1
## factor(seats)5                   668.2793    1
## factor(seats)6                   32.4156     1
## factor(seats)7                   489.0089    1
## factor(seats)8                   116.7583    1
## factor(seats)9                   41.1450     1
## factor(seats)10                  10.6120     1
## factor(seats)14                  1.5064      0
##
## Multicollinearity may be due to factor(seats)4 factor(seats)5 factor(seats)6 factor(seats)7
## 
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
## 
## =====

```

From the pairs plot and vif calculation above, there is no correlation between any of the predictor variables. Hence, we conclude that there is no multi-collinearity among the predictor variables. However, there is a correlation between the seat predictor variable itself.

### 3. Implement the full model

```
#Fit a linear model based on selected predictor variables.

firstordermodel = lm(selling_price ~ age + km_driven + factor(fuel) +
  ↴ factor(seller_type)
  + factor(transmission)+ factor(owner)+ mileage +
  ↴ engine + max_power + factor(seats)
  ,data = Car1)
```

```
summary(firstordermodel)
```

```
## 
## Call:
## lm(formula = selling_price ~ age + km_driven + factor(fuel) +
```

```

##      factor(seller_type) + factor(transmission) + factor(owner) +
##      mileage + engine + max_power + factor(seats), data = Car1)
##
## Residuals:
##   Min     1Q Median     3Q    Max
## -37118 -3351    165   2645  62928
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                4.700e+03  5.413e+03   0.868 0.385296
## age                      -6.641e+02  3.267e+01  -20.329 < 2e-16 ***
## km_driven                  -1.533e-02  1.805e-03  -8.490 < 2e-16 ***
## factor(fuel)Petrol        -8.852e+02  2.519e+02  -3.514 0.000444 ***
## factor(seller_type)Individual -4.010e+03  2.717e+02  -14.760 < 2e-16 ***
## factor(seller_type)Trustmark Dealer -5.514e+03  5.557e+02  -9.922 < 2e-16 ***
## factor(transmission)Manual    -7.085e+03  3.271e+02  -21.658 < 2e-16 ***
## factor(owner)Fourth & Above Owner 2.405e+02  6.348e+02   0.379 0.704834
## factor(owner)Second Owner    -5.281e+02  2.213e+02  -2.386 0.017040 *
## factor(owner)Test Drive Car   3.500e+04  3.377e+03  10.363 < 2e-16 ***
## factor(owner)Third Owner     -7.201e+01  3.802e+02  -0.189 0.849790
## mileage                     2.253e+02  3.649e+01   6.175 6.96e-10 ***
## engine                       1.466e+00  3.969e-01   3.695 0.000222 ***
## max_power                    2.193e+02  4.367e+00  50.231 < 2e-16 ***
## factor(seats)4               6.300e+03  5.448e+03   1.156 0.247556
## factor(seats)5               -5.234e+03  5.395e+03  -0.970 0.331931
## factor(seats)6               -7.420e+03  5.462e+03  -1.358 0.174381
## factor(seats)7               -6.996e+03  5.373e+03  -1.302 0.192963
## factor(seats)8               -5.114e+03  5.385e+03  -0.950 0.342339
## factor(seats)9               -3.981e+03  5.424e+03  -0.734 0.462936
## factor(seats)10              -2.457e+03  5.630e+03  -0.436 0.662488
## factor(seats)14              -7.291e+03  9.236e+03  -0.789 0.429845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7524 on 7797 degrees of freedom
## Multiple R-squared:  0.6955, Adjusted R-squared:  0.6947
## F-statistic: 848.1 on 21 and 7797 DF,  p-value: < 2.2e-16

```

From the result of fitting a linear model using all possible independent variables, the  $R^2_{adj}$  is: 0.6947; meaning 69.47% of the variation in selling price of a used car, can be explained by the model.

#### 4. Use stepwise regression (with backward elimination) to find the ‘best’ set of predictors of selling\_price

```

# stepwise
library(olsrr)

## 
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
## 
##      rivers

stepmod=ols_step_backward_p(firstordermodel, prem = 0.3, details=TRUE)

## Backward Elimination Method
## -----
## 
## Candidate Terms:
## 
## 1 . age
## 2 . km_driven
## 3 . factor(fuel)
## 4 . factor(seller_type)
## 5 . factor(transmission)
## 6 . factor(owner)
## 7 . mileage
## 8 . engine
## 9 . max_power
## 10 . factor(seats)
## 
## We are eliminating variables based on p value...
## 
## 
## No more variables satisfy the condition of p value = 0.3
## 
## 
## Variables Removed:
## 
## 
## 
## Final Model Output
## -----
## 
##          Model Summary

```

```

## -----
## R 0.834 RMSE 7524.223
## R-Squared 0.696 Coef. Var 69.031
## Adj. R-Squared 0.695 MSE 56613939.167
## Pred R-Squared -Inf MAE 4582.265
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
## ANOVA
## -----
## Sum of
## Squares DF Mean Square F Sig.
## -----
## Regression 1.008332e+12 21 48015828002.501 848.127 0.0000
## Residual 441418883682.351 7797 56613939.167
## Total 1.449751e+12 7818
## -----
## Parameter Estimates
## -----
## model Beta Std. Error Std. Beta t
## -----
## (Intercept) 4699.532 5412.779 0.86
## age -664.082 32.667 -0.188 -20.32
## km_driven -0.015 0.002 -0.064 -8.49
## factor(fuel)Petrol -885.179 251.908 -0.032 -3.51
## factor(seller_type)Individual -4009.808 271.667 -0.111 -14.76
## factor(seller_type)Trustmark Dealer -5513.629 555.682 -0.069 -9.92
## factor(transmission)Manual -7084.544 327.117 -0.177 -21.65
## factor(owner)Fourth & Above Owner 240.454 634.751 0.002 0.37
## factor(owner)Second Owner -528.065 221.285 -0.017 -2.38
## factor(owner)Test Drive Car 34996.846 3377.230 0.065 10.36
## factor(owner)Third Owner -72.013 380.231 -0.001 -0.18
## mileage 225.326 36.492 0.066 6.17
## engine 1.466 0.397 0.054 3.69
## max_power 219.338 4.367 0.576 50.23
## factor(seats)4 6300.189 5448.139 0.059 1.15
## factor(seats)5 -5234.306 5394.566 -0.157 -0.97
## factor(seats)6 -7419.828 5462.247 -0.048 -1.35
## factor(seats)7 -6995.941 5373.327 -0.180 -1.30
## factor(seats)8 -5113.804 5385.153 -0.064 -0.95
## factor(seats)9 -3981.425 5423.860 -0.029 -0.73
## factor(seats)10 -2457.500 5630.050 -0.009 -0.43

```

```

##                               factor(seats)14      -7291.412      9235.505      -0.006      -0.78
## -----
## 
summary(stepmod$model)

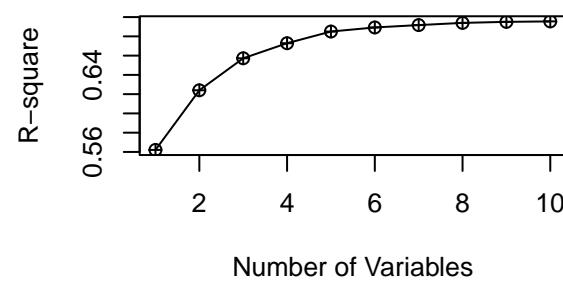
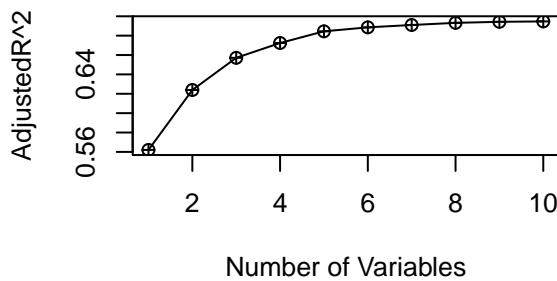
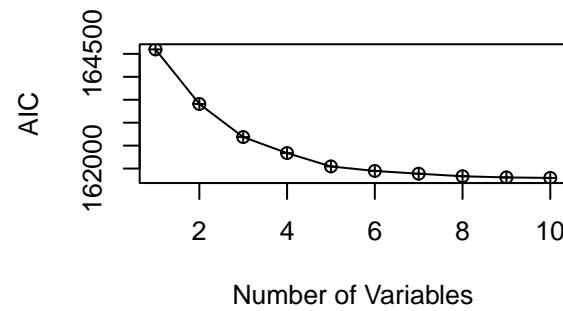
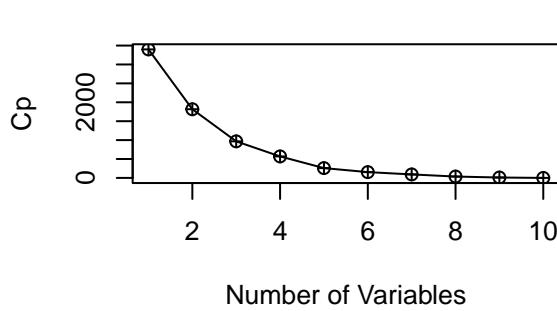
## 
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = 1)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -37118 -3351    165   2645  62928 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)          4.700e+03  5.413e+03   0.868 0.385296  
## age                  -6.641e+02  3.267e+01  -20.329 < 2e-16 ***
## km_driven            -1.533e-02  1.805e-03  -8.490 < 2e-16 *** 
## factor(fuel)Petrol   -8.852e+02  2.519e+02  -3.514 0.000444 *** 
## factor(seller_type)Individual -4.010e+03  2.717e+02  -14.760 < 2e-16 *** 
## factor(seller_type)Trustmark Dealer -5.514e+03  5.557e+02  -9.922 < 2e-16 *** 
## factor(transmission)Manual   -7.085e+03  3.271e+02  -21.658 < 2e-16 *** 
## factor(owner)Fourth & Above Owner  2.405e+02  6.348e+02   0.379 0.704834  
## factor(owner)Second Owner       -5.281e+02  2.213e+02  -2.386 0.017040 *  
## factor(owner)Test Drive Car    3.500e+04  3.377e+03  10.363 < 2e-16 *** 
## factor(owner)Third Owner      -7.201e+01  3.802e+02  -0.189 0.849790  
## mileage                2.253e+02  3.649e+01   6.175 6.96e-10 *** 
## engine                 1.466e+00  3.969e-01   3.695 0.000222 *** 
## max_power              2.193e+02  4.367e+00  50.231 < 2e-16 *** 
## factor(seats)4           6.300e+03  5.448e+03   1.156 0.247556  
## factor(seats)5           -5.234e+03  5.395e+03  -0.970 0.331931  
## factor(seats)6           -7.420e+03  5.462e+03  -1.358 0.174381  
## factor(seats)7           -6.996e+03  5.373e+03  -1.302 0.192963  
## factor(seats)8           -5.114e+03  5.385e+03  -0.950 0.342339  
## factor(seats)9           -3.981e+03  5.424e+03  -0.734 0.462936  
## factor(seats)10          -2.457e+03  5.630e+03  -0.436 0.662488  
## factor(seats)14          -7.291e+03  9.236e+03  -0.789 0.429845  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 7524 on 7797 degrees of freedom
## Multiple R-squared:  0.6955, Adjusted R-squared:  0.6947 
## F-statistic: 848.1 on 21 and 7797 DF,  p-value: < 2.2e-16

```

From the result of the backwards elimination model, suitable significant predictors of the selling price of a used car are: age, km\_driven, fuel, seller\_type, transmission, owner, mileage, engine and max\_power

##### 5. Use all-possible-regressions-selection to find the ‘best’ predictors of selling\_price (Cp, AIC, Adjusted R<sup>2</sup>)

```
library(olsrr)
best_model = ols_step_best_subset(firstordermodel, details = TRUE)
par(mfrow=c(2,2))
plot(best_model$cp,type = "o",pch = 10, xlab= "Number of Variables",
     ylab = "Cp")
plot(best_model$aic,type = "o",pch = 10, xlab= "Number of Variables",
     ylab = "AIC")
plot(best_model$adjr,type = "o",pch = 10, xlab= "Number of Variables",
     ylab = "AdjustedR^2")
plot(best_model$rsq,type = "o",pch = 10, xlab= "Number of Variables",
     ylab = "R-square")
```



From the output above, the model with 9 predictors has the lowest Cp and AIC. Furthermore, it also has the highest  $R_{adj}^2$  value. Hence, the model with 9 predictors will be the best at predicting the selling price of a used car.

## 6. Test the hypothesis for the full model.

Using the Global F test to confirm that the independent variable are suitable for predicting the selling price of a used car at significance level 0.05.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_a : \text{at least one of } \beta_i \text{ in not zero (i= 1,2,..p)}$$

```
#If the p-value of f-test is less than 0.05, conduct a partial t-test to
# check the significance of each predictor variable.
new_firstrdermodel = lm(selling_price ~ age + km_driven + factor(fuel) +
# factor(seller_type)
# + factor(transmission)+ factor(owner)+ mileage +
# engine + max_power
,data = Car1)
summary(new_firstrdermodel)
```

```
##
## Call:
## lm(formula = selling_price ~ age + km_driven + factor(fuel) +
##     factor(seller_type) + factor(transmission) + factor(owner) +
##     mileage + engine + max_power, data = Car1)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -37119   -3256    161   2493  74403
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.782e+02  1.292e+03  0.138  0.89026
## age                      -5.096e+02  3.124e+01 -16.312 < 2e-16 ***
## km_driven                  -1.803e-02  1.828e-03 -9.864 < 2e-16 ***
## factor(fuel)Petrol          -7.116e+02  2.529e+02 -2.814  0.00491 **
## factor(seller_type)Individual -4.165e+03  2.759e+02 -15.095 < 2e-16 ***
## factor(seller_type)Trustmark Dealer -5.780e+03  5.652e+02 -10.226 < 2e-16 ***
## factor(transmission)Manual     -7.842e+03  3.272e+02 -23.970 < 2e-16 ***
## factor(owner)Fourth & Above Owner 1.745e+02  6.465e+02  0.270  0.78718
## factor(owner)Second Owner     -7.138e+02  2.250e+02 -3.172  0.00152 **
## factor(owner)Test Drive Car    3.528e+04  3.441e+03 10.253 < 2e-16 ***
## factor(owner)Third Owner      -3.111e+02  3.870e+02 -0.804  0.42156
## mileage                      2.495e+02  3.502e+01  7.124 1.14e-12 ***
## engine                        7.510e-01  3.596e-01  2.089  0.03678 *
## max_power                     2.177e+02  4.093e+00 53.189 < 2e-16 ***
## ---
```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7668 on 7805 degrees of freedom
## Multiple R-squared:  0.6835, Adjusted R-squared:  0.683
## F-statistic:  1296 on 13 and 7805 DF,  p-value: < 2.2e-16

# Conduct a global F-test to check for overall model significance.
reduced_reg = lm(selling_price ~ 1,data = Car1)
anova(reduced_reg,new_firstordermodel)

## Analysis of Variance Table
##
## Model 1: selling_price ~ 1
## Model 2: selling_price ~ age + km_driven + factor(fuel) + factor(seller_type) +
##           factor(transmission) + factor(owner) + mileage + engine +
##           max_power
##   Res.Df      RSS Df  Sum of Sq      F    Pr(>F)
## 1    7818 1.4498e+12
## 2    7805 4.5886e+11 13 9.9089e+11 1296.5 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

From the result of the full model F-test conducted above, the F-calc is 1296 and the p-value: 2.2e-16 < 0.05. This indicates that atleast one of the independent predictor variable is not equal to 0.

## 7. Conducting a partial F-test, to test if number of seats should be kept in the model.

$$H_0 : \beta_{10} = 0$$

$$H_a : \beta_{10} \neq 0$$

```

# partial F-test
anova(new_firstordermodel,firstordermodel)

```

```

## Analysis of Variance Table
##
## Model 1: selling_price ~ age + km_driven + factor(fuel) + factor(seller_type) +
##           factor(transmission) + factor(owner) + mileage + engine +
##           max_power
## Model 2: selling_price ~ age + km_driven + factor(fuel) + factor(seller_type) +
##           factor(transmission) + factor(owner) + mileage + engine +
##           max_power + factor(seats)

```

```

##   Res.Df      RSS Df  Sum of Sq      F    Pr(>F)
## 1    7805 4.5886e+11
## 2    7797 4.4142e+11  8 1.7444e+10 38.514 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

From the result of the partial F-test above, since F-calc is 38.514 and the p-value: 2.2e-16 < 0.05, we reject the null hypothesis. This indicates that the seat predictor should be kept in the model.

Although the result of the partial F-test conducted above rejects the null hypothesis, indicating the number of seats variable be kept in the model, but by conducting the backward elimination, the best subset and a partial t-test, we concluded on using the model without the seat predictor variable.

## 8. Build an interaction model to fit the multiple regression model.

*#Check for interaction terms among the significant predictor variables.*

```

Interac_firstordermodel = lm(selling_price ~ (age + km_driven + factor(fuel)
  ↵ + factor(seller_type)
  ↵ + factor(transmission)+ factor(owner)+ mileage +
  ↵ engine + max_power)^2
  ,data = Car1)
summary(Interac_firstordermodel)

```

```

##
## Call:
## lm(formula = selling_price ~ (age + km_driven + factor(fuel) +
##     factor(seller_type) + factor(transmission) + factor(owner) +
##     mileage + engine + max_power)^2, data = Car1)
##
## Residuals:
##    Min      1Q  Median      3Q      Max
## -44253  -1419     15  1411  35330
##
## Coefficients: (9 not defined because of singularities)
##                                     Estimate
## (Intercept)                      -5.449e+03
## age                                -1.170e+03
## km_driven                           -6.463e-02
## factor(fuel)Petrol                  -1.654e+04
## factor(seller_type)Individual       1.010e+04
## factor(seller_type)Trustmark Dealer -5.235e+03
## factor(transmission)Manual          3.279e+02

```

## factor(owner)Fourth & Above Owner	8.003e+03
## factor(owner)Second Owner	-2.124e+02
## factor(owner)Test Drive Car	9.126e+05
## factor(owner)Third Owner	1.715e+04
## mileage	1.178e+02
## engine	-1.242e+01
## max_power	4.514e+02
## age:km_driven	2.270e-03
## age:factor(fuel)Petrol	5.630e+02
## age:factor(seller_type)Individual	4.053e+02
## age:factor(seller_type)Trustmark Dealer	-1.281e+02
## age:factor(transmission)Manual	8.996e+02
## age:factor(owner)Fourth & Above Owner	2.059e+02
## age:factor(owner)Second Owner	7.829e+01
## age:factor(owner)Test Drive Car	NA
## age:factor(owner)Third Owner	-2.089e+01
## age:mileage	-7.159e+00
## age:engine	3.145e-01
## age:max_power	-2.070e+01
## km_driven:factor(fuel)Petrol	-2.120e-03
## km_driven:factor(seller_type)Individual	2.162e-02
## km_driven:factor(seller_type)Trustmark Dealer	9.690e-02
## km_driven:factor(transmission)Manual	4.362e-02
## km_driven:factor(owner)Fourth & Above Owner	1.355e-03
## km_driven:factor(owner)Second Owner	-4.624e-03
## km_driven:factor(owner)Test Drive Car	5.541e-01
## km_driven:factor(owner)Third Owner	-6.111e-04
## km_driven:mileage	-4.358e-05
## km_driven:engine	7.052e-06
## km_driven:max_power	-3.125e-04
## factor(fuel)Petrol:factor(seller_type)Individual	-5.694e+02
## factor(fuel)Petrol:factor(seller_type)Trustmark Dealer	8.612e+02
## factor(fuel)Petrol:factor(transmission)Manual	3.139e+03
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	-2.666e+03
## factor(fuel)Petrol:factor(owner)Second Owner	-3.381e+02
## factor(fuel)Petrol:factor(owner)Test Drive Car	-2.504e+05
## factor(fuel)Petrol:factor(owner)Third Owner	-1.521e+03
## factor(fuel)Petrol:mileage	3.047e+02
## factor(fuel)Petrol:engine	6.554e+00
## factor(fuel)Petrol:max_power	-5.176e+01
## factor(seller_type)Individual:factor(transmission)Manual	-4.791e+02
## factor(seller_type)Trustmark Dealer:factor(transmission)Manual	1.780e+03
## factor(seller_type)Individual:factor(owner)Fourth & Above Owner	NA
## factor(seller_type)Trustmark Dealer:factor(owner)Fourth & Above Owner	NA
## factor(seller_type)Individual:factor(owner)Second Owner	-5.381e+02

## factor(seller_type)Trustmark Dealer:factor(owner)Second Owner	-2.900e+03
## factor(seller_type)Individual:factor(owner)Test Drive Car	NA
## factor(seller_type)Trustmark Dealer:factor(owner)Test Drive Car	NA
## factor(seller_type)Individual:factor(owner)Third Owner	-3.920e+03
## factor(seller_type)Trustmark Dealer:factor(owner)Third Owner	NA
## factor(seller_type)Individual:mileage	-2.660e+02
## factor(seller_type)Trustmark Dealer:mileage	-4.529e+01
## factor(seller_type)Individual:engine	1.833e+00
## factor(seller_type)Trustmark Dealer:engine	5.560e+00
## factor(seller_type)Individual:max_power	-1.086e+02
## factor(seller_type)Trustmark Dealer:max_power	-8.121e+01
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	-5.346e+03
## factor(transmission)Manual:factor(owner)Second Owner	-7.262e+02
## factor(transmission)Manual:factor(owner)Test Drive Car	NA
## factor(transmission)Manual:factor(owner)Third Owner	-5.643e+03
## factor(transmission)Manual:mileage	3.976e+01
## factor(transmission)Manual:engine	1.164e+01
## factor(transmission)Manual:max_power	-2.627e+02
## factor(owner)Fourth & Above Owner:mileage	-1.212e+02
## factor(owner)Second Owner:mileage	5.584e+01
## factor(owner)Test Drive Car:mileage	-4.120e+04
## factor(owner)Third Owner:mileage	-2.421e+02
## factor(owner)Fourth & Above Owner:engine	-3.251e+00
## factor(owner)Second Owner:engine	-6.467e-02
## factor(owner)Test Drive Car:engine	NA
## factor(owner)Third Owner:engine	-2.888e+00
## factor(owner)Fourth & Above Owner:max_power	2.934e+01
## factor(owner)Second Owner:max_power	2.221e+00
## factor(owner)Test Drive Car:max_power	NA
## factor(owner)Third Owner:max_power	1.718e+01
## mileage:engine	-2.822e-01
## mileage:max_power	5.558e+00
## engine:max_power	3.219e-02
##	Std. Error
## (Intercept)	4.299e+03
## age	2.237e+02
## km_driven	1.991e-02
## factor(fuel)Petrol	1.764e+03
## factor(seller_type)Individual	2.099e+03
## factor(seller_type)Trustmark Dealer	5.662e+03
## factor(transmission)Manual	2.896e+03
## factor(owner)Fourth & Above Owner	5.207e+03
## factor(owner)Second Owner	2.038e+03
## factor(owner)Test Drive Car	1.060e+05
## factor(owner)Third Owner	4.077e+03

## mileage	1.306e+02
## engine	1.634e+00
## max_power	1.942e+01
## age:km_driven	3.439e-04
## age:factor(fuel)Petrol	5.496e+01
## age:factor(seller_type)Individual	6.777e+01
## age:factor(seller_type)Trustmark Dealer	3.693e+02
## age:factor(transmission)Manual	8.261e+01
## age:factor(owner)Fourth & Above Owner	1.025e+02
## age:factor(owner)Second Owner	3.959e+01
## age:factor(owner)Test Drive Car	NA
## age:factor(owner)Third Owner	6.842e+01
## age:mileage	6.025e+00
## age:engine	7.051e-02
## age:max_power	8.072e-01
## km_driven:factor(fuel)Petrol	3.663e-03
## km_driven:factor(seller_type)Individual	5.715e-03
## km_driven:factor(seller_type)Trustmark Dealer	3.110e-02
## km_driven:factor(transmission)Manual	6.217e-03
## km_driven:factor(owner)Fourth & Above Owner	8.320e-03
## km_driven:factor(owner)Second Owner	2.717e-03
## km_driven:factor(owner)Test Drive Car	3.468e-01
## km_driven:factor(owner)Third Owner	4.894e-03
## km_driven:mileage	5.548e-04
## km_driven:engine	4.787e-06
## km_driven:max_power	4.789e-05
## factor(fuel)Petrol:factor(seller_type)Individual	4.491e+02
## factor(fuel)Petrol:factor(seller_type)Trustmark Dealer	2.268e+03
## factor(fuel)Petrol:factor(transmission)Manual	6.037e+02
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	1.078e+03
## factor(fuel)Petrol:factor(owner)Second Owner	3.817e+02
## factor(fuel)Petrol:factor(owner)Test Drive Car	3.509e+04
## factor(fuel)Petrol:factor(owner)Third Owner	6.222e+02
## factor(fuel)Petrol:mileage	4.794e+01
## factor(fuel)Petrol:engine	8.534e-01
## factor(fuel)Petrol:max_power	9.123e+00
## factor(seller_type)Individual:factor(transmission)Manual	4.479e+02
## factor(seller_type)Trustmark Dealer:factor(transmission)Manual	1.034e+03
## factor(seller_type)Individual:factor(owner)Fourth & Above Owner	NA
## factor(seller_type)Trustmark Dealer:factor(owner)Fourth & Above Owner	NA
## factor(seller_type)Individual:factor(owner)Second Owner	5.065e+02
## factor(seller_type)Trustmark Dealer:factor(owner)Second Owner	2.523e+03
## factor(seller_type)Individual:factor(owner)Test Drive Car	NA
## factor(seller_type)Trustmark Dealer:factor(owner)Test Drive Car	NA
## factor(seller_type)Individual:factor(owner)Third Owner	2.075e+03

## factor(seller_type)Trustmark Dealer:factor(owner)Third Owner	NA
## factor(seller_type)Individual:mileage	5.796e+01
## factor(seller_type)Trustmark Dealer:mileage	1.767e+02
## factor(seller_type)Individual:engine	7.175e-01
## factor(seller_type)Trustmark Dealer:engine	2.227e+00
## factor(seller_type)Individual:max_power	7.379e+00
## factor(seller_type)Trustmark Dealer:max_power	4.038e+01
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	1.890e+03
## factor(transmission)Manual:factor(owner)Second Owner	5.487e+02
## factor(transmission)Manual:factor(owner)Test Drive Car	NA
## factor(transmission)Manual:factor(owner)Third Owner	1.144e+03
## factor(transmission)Manual:mileage	8.038e+01
## factor(transmission)Manual:engine	8.760e-01
## factor(transmission)Manual:max_power	7.432e+00
## factor(owner)Fourth & Above Owner:mileage	1.435e+02
## factor(owner)Second Owner:mileage	5.238e+01
## factor(owner)Test Drive Car:mileage	4.808e+03
## factor(owner)Third Owner:mileage	9.820e+01
## factor(owner)Fourth & Above Owner:engine	1.398e+00
## factor(owner)Second Owner:engine	5.047e-01
## factor(owner)Test Drive Car:engine	NA
## factor(owner)Third Owner:engine	9.131e-01
## factor(owner)Fourth & Above Owner:max_power	1.833e+01
## factor(owner)Second Owner:max_power	5.946e+00
## factor(owner)Test Drive Car:max_power	NA
## factor(owner)Third Owner:max_power	1.039e+01
## mileage:engine	5.604e-02
## mileage:max_power	5.887e-01
## engine:max_power	4.922e-03
##	t value
## (Intercept)	-1.267
## age	-5.233
## km_driven	-3.247
## factor(fuel)Petrol	-9.377
## factor(seller_type)Individual	4.813
## factor(seller_type)Trustmark Dealer	-0.925
## factor(transmission)Manual	0.113
## factor(owner)Fourth & Above Owner	1.537
## factor(owner)Second Owner	-0.104
## factor(owner)Test Drive Car	8.608
## factor(owner)Third Owner	4.206
## mileage	0.902
## engine	-7.605
## max_power	23.246
## age:km_driven	6.602

## age:factor(fuel)Petrol	10.244
## age:factor(seller_type)Individual	5.981
## age:factor(seller_type)Trustmark Dealer	-0.347
## age:factor(transmission)Manual	10.890
## age:factor(owner)Fourth & Above Owner	2.009
## age:factor(owner)Second Owner	1.978
## age:factor(owner)Test Drive Car	NA
## age:factor(owner)Third Owner	-0.305
## age:mileage	-1.188
## age:engine	4.460
## age:max_power	-25.651
## km_driven:factor(fuel)Petrol	-0.579
## km_driven:factor(seller_type)Individual	3.783
## km_driven:factor(seller_type)Trustmark Dealer	3.116
## km_driven:factor(transmission)Manual	7.016
## km_driven:factor(owner)Fourth & Above Owner	0.163
## km_driven:factor(owner)Second Owner	-1.702
## km_driven:factor(owner)Test Drive Car	1.598
## km_driven:factor(owner)Third Owner	-0.125
## km_driven:mileage	-0.079
## km_driven:engine	1.473
## km_driven:max_power	-6.525
## factor(fuel)Petrol:factor(seller_type)Individual	-1.268
## factor(fuel)Petrol:factor(seller_type)Trustmark Dealer	0.380
## factor(fuel)Petrol:factor(transmission)Manual	5.199
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	-2.473
## factor(fuel)Petrol:factor(owner)Second Owner	-0.886
## factor(fuel)Petrol:factor(owner)Test Drive Car	-7.136
## factor(fuel)Petrol:factor(owner)Third Owner	-2.444
## factor(fuel)Petrol:mileage	6.355
## factor(fuel)Petrol:engine	7.680
## factor(fuel)Petrol:max_power	-5.673
## factor(seller_type)Individual:factor(transmission)Manual	-1.070
## factor(seller_type)Trustmark Dealer:factor(transmission)Manual	1.721
## factor(seller_type)Individual:factor(owner)Fourth & Above Owner	NA
## factor(seller_type)Trustmark Dealer:factor(owner)Fourth & Above Owner	NA
## factor(seller_type)Individual:factor(owner)Second Owner	-1.062
## factor(seller_type)Trustmark Dealer:factor(owner)Second Owner	-1.150
## factor(seller_type)Individual:factor(owner)Test Drive Car	NA
## factor(seller_type)Trustmark Dealer:factor(owner)Test Drive Car	NA
## factor(seller_type)Individual:factor(owner)Third Owner	-1.889
## factor(seller_type)Trustmark Dealer:factor(owner)Third Owner	NA
## factor(seller_type)Individual:mileage	-4.590
## factor(seller_type)Trustmark Dealer:mileage	-0.256
## factor(seller_type)Individual:engine	2.555

## factor(seller_type)Trustmark Dealer:engine	2.497
## factor(seller_type)Individual:max_power	-14.719
## factor(seller_type)Trustmark Dealer:max_power	-2.011
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	-2.829
## factor(transmission)Manual:factor(owner)Second Owner	-1.323
## factor(transmission)Manual:factor(owner)Test Drive Car	NA
## factor(transmission)Manual:factor(owner)Third Owner	-4.931
## factor(transmission)Manual:mileage	0.495
## factor(transmission)Manual:engine	13.292
## factor(transmission)Manual:max_power	-35.352
## factor(owner)Fourth & Above Owner:mileage	-0.844
## factor(owner)Second Owner:mileage	1.066
## factor(owner)Test Drive Car:mileage	-8.569
## factor(owner)Third Owner:mileage	-2.466
## factor(owner)Fourth & Above Owner:engine	-2.325
## factor(owner)Second Owner:engine	-0.128
## factor(owner)Test Drive Car:engine	NA
## factor(owner)Third Owner:engine	-3.163
## factor(owner)Fourth & Above Owner:max_power	1.601
## factor(owner)Second Owner:max_power	0.373
## factor(owner)Test Drive Car:max_power	NA
## factor(owner)Third Owner:max_power	1.653
## mileage:engine	-5.035
## mileage:max_power	9.440
## engine:max_power	6.539
##	Pr(> t )
## (Intercept)	0.205019
## age	1.71e-07
## km_driven	0.001171
## factor(fuel)Petrol	< 2e-16
## factor(seller_type)Individual	1.52e-06
## factor(seller_type)Trustmark Dealer	0.355146
## factor(transmission)Manual	0.909847
## factor(owner)Fourth & Above Owner	0.124312
## factor(owner)Second Owner	0.917022
## factor(owner)Test Drive Car	< 2e-16
## factor(owner)Third Owner	2.63e-05
## mileage	0.367015
## engine	3.19e-14
## max_power	< 2e-16
## age:km_driven	4.33e-11
## age:factor(fuel)Petrol	< 2e-16
## age:factor(seller_type)Individual	2.31e-09
## age:factor(seller_type)Trustmark Dealer	0.728580
## age:factor(transmission)Manual	< 2e-16

## age:factor(owner)Fourth & Above Owner	0.044544
## age:factor(owner)Second Owner	0.048007
## age:factor(owner)Test Drive Car	NA
## age:factor(owner)Third Owner	0.760168
## age:mileage	0.234808
## age:engine	8.32e-06
## age:max_power	< 2e-16
## km_driven:factor(fuel)Petrol	0.562737
## km_driven:factor(seller_type)Individual	0.000156
## km_driven:factor(seller_type)Trustmark Dealer	0.001841
## km_driven:factor(transmission)Manual	2.47e-12
## km_driven:factor(owner)Fourth & Above Owner	0.870610
## km_driven:factor(owner)Second Owner	0.088838
## km_driven:factor(owner)Test Drive Car	0.110157
## km_driven:factor(owner)Third Owner	0.900622
## km_driven:mileage	0.937387
## km_driven:engine	0.140768
## km_driven:max_power	7.21e-11
## factor(fuel)Petrol:factor(seller_type)Individual	0.204814
## factor(fuel)Petrol:factor(seller_type)Trustmark Dealer	0.704124
## factor(fuel)Petrol:factor(transmission)Manual	2.06e-07
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	0.013429
## factor(fuel)Petrol:factor(owner)Second Owner	0.375750
## factor(fuel)Petrol:factor(owner)Test Drive Car	1.05e-12
## factor(fuel)Petrol:factor(owner)Third Owner	0.014548
## factor(fuel)Petrol:mileage	2.20e-10
## factor(fuel)Petrol:engine	1.79e-14
## factor(fuel)Petrol:max_power	1.45e-08
## factor(seller_type)Individual:factor(transmission)Manual	0.284807
## factor(seller_type)Trustmark Dealer:factor(transmission)Manual	0.085268
## factor(seller_type)Individual:factor(owner)Fourth & Above Owner	NA
## factor(seller_type)Trustmark Dealer:factor(owner)Fourth & Above Owner	NA
## factor(seller_type)Individual:factor(owner)Second Owner	0.288127
## factor(seller_type)Trustmark Dealer:factor(owner)Second Owner	0.250291
## factor(seller_type)Individual:factor(owner)Test Drive Car	NA
## factor(seller_type)Trustmark Dealer:factor(owner)Test Drive Car	NA
## factor(seller_type)Individual:factor(owner)Third Owner	0.058899
## factor(seller_type)Trustmark Dealer:factor(owner)Third Owner	NA
## factor(seller_type)Individual:mileage	4.51e-06
## factor(seller_type)Trustmark Dealer:mileage	0.797724
## factor(seller_type)Individual:engine	0.010632
## factor(seller_type)Trustmark Dealer:engine	0.012537
## factor(seller_type)Individual:max_power	< 2e-16
## factor(seller_type)Trustmark Dealer:max_power	0.044333
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	0.004686

## factor(transmission)Manual:factor(owner)Second Owner	0.185730
## factor(transmission)Manual:factor(owner)Test Drive Car	NA
## factor(transmission)Manual:factor(owner)Third Owner	8.37e-07
## factor(transmission)Manual:mileage	0.620878
## factor(transmission)Manual:engine	< 2e-16
## factor(transmission)Manual:max_power	< 2e-16
## factor(owner)Fourth & Above Owner:mileage	0.398505
## factor(owner)Second Owner:mileage	0.286426
## factor(owner)Test Drive Car:mileage	< 2e-16
## factor(owner)Third Owner:mileage	0.013696
## factor(owner)Fourth & Above Owner:engine	0.020072
## factor(owner)Second Owner:engine	0.898042
## factor(owner)Test Drive Car:engine	NA
## factor(owner)Third Owner:engine	0.001569
## factor(owner)Fourth & Above Owner:max_power	0.109451
## factor(owner)Second Owner:max_power	0.708803
## factor(owner)Test Drive Car:max_power	NA
## factor(owner)Third Owner:max_power	0.098317
## mileage:engine	4.88e-07
## mileage:max_power	< 2e-16
## engine:max_power	6.58e-11
##	
## (Intercept)	
## age	***
## km_driven	**
## factor(fuel)Petrol	***
## factor(seller_type)Individual	***
## factor(seller_type)Trustmark Dealer	***
## factor(transmission)Manual	
## factor(owner)Fourth & Above Owner	
## factor(owner)Second Owner	
## factor(owner)Test Drive Car	***
## factor(owner)Third Owner	***
## mileage	
## engine	***
## max_power	***
## age:km_driven	***
## age:factor(fuel)Petrol	***
## age:factor(seller_type)Individual	***
## age:factor(seller_type)Trustmark Dealer	
## age:factor(transmission)Manual	***
## age:factor(owner)Fourth & Above Owner	*
## age:factor(owner)Second Owner	*
## age:factor(owner)Test Drive Car	
## age:factor(owner)Third Owner	

```

## age:mileage
## age:engine ***

## age:max_power ***

## km_driven:factor(fuel)Petrol
## km_driven:factor(seller_type)Individual ***
## km_driven:factor(seller_type)Trustmark Dealer **
## km_driven:factor(transmission)Manual ***
## km_driven:factor(owner)Fourth & Above Owner
## km_driven:factor(owner)Second Owner .
## km_driven:factor(owner)Test Drive Car
## km_driven:factor(owner)Third Owner

## km_driven:mileage
## km_driven:engine
## km_driven:max_power ***

## factor(fuel)Petrol:factor(seller_type)Individual
## factor(fuel)Petrol:factor(seller_type)Trustmark Dealer
## factor(fuel)Petrol:factor(transmission)Manual ***
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner *
## factor(fuel)Petrol:factor(owner)Second Owner
## factor(fuel)Petrol:factor(owner)Test Drive Car ***
## factor(fuel)Petrol:factor(owner)Third Owner *
## factor(fuel)Petrol:mileage ***
## factor(fuel)Petrol:engine ***
## factor(fuel)Petrol:max_power ***

## factor(seller_type)Individual:factor(transmission)Manual
## factor(seller_type)Trustmark Dealer:factor(transmission)Manual .
## factor(seller_type)Individual:factor(owner)Fourth & Above Owner
## factor(seller_type)Trustmark Dealer:factor(owner)Fourth & Above Owner
## factor(seller_type)Individual:factor(owner)Second Owner
## factor(seller_type)Trustmark Dealer:factor(owner)Second Owner
## factor(seller_type)Individual:factor(owner)Test Drive Car
## factor(seller_type)Trustmark Dealer:factor(owner)Test Drive Car
## factor(seller_type)Individual:factor(owner)Third Owner .
## factor(seller_type)Trustmark Dealer:factor(owner)Third Owner
## factor(seller_type)Individual:mileage ***
## factor(seller_type)Trustmark Dealer:mileage
## factor(seller_type)Individual:engine *
## factor(seller_type)Trustmark Dealer:engine *
## factor(seller_type)Individual:max_power ***
## factor(seller_type)Trustmark Dealer:max_power *
## factor(transmission)Manual:factor(owner)Fourth & Above Owner **
## factor(transmission)Manual:factor(owner)Second Owner
## factor(transmission)Manual:factor(owner)Test Drive Car
## factor(transmission)Manual:factor(owner)Third Owner ***
## factor(transmission)Manual:mileage

```

```

## factor(transmission)Manual:engine ***  

## factor(transmission)Manual:max_power ***  

## factor(owner)Fourth & Above Owner:mileage  

## factor(owner)Second Owner:mileage  

## factor(owner)Test Drive Car:mileage ***  

## factor(owner)Third Owner:mileage *  

## factor(owner)Fourth & Above Owner:engine *  

## factor(owner)Second Owner:engine  

## factor(owner)Test Drive Car:engine  

## factor(owner)Third Owner:engine **  

## factor(owner)Fourth & Above Owner:max_power  

## factor(owner)Second Owner:max_power  

## factor(owner)Test Drive Car:max_power  

## factor(owner)Third Owner:max_power .  

## mileage:engine ***  

## mileage:max_power ***  

## engine:max_power ***  

## ---  

## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  

##  

## Residual standard error: 4052 on 7743 degrees of freedom  

## Multiple R-squared: 0.9123, Adjusted R-squared: 0.9114  

## F-statistic: 1074 on 75 and 7743 DF, p-value: < 2.2e-16

```

From the result of fitting interaction terms above, since not all the interaction terms are significant, we will refit the model with only the significant interaction terms.

## 9. Conduct a partial t-test to check for significance of interaction term.

$$H_0 : \beta_i = 0$$

$$H_a : \beta_i \neq 0$$

```

# Interaction model  

#Conduct a partial t-test to check for significance of interaction term.  

new_Interac_firstordermodel = lm(selling_price ~ age + km_driven +  

  ↵ factor(fuel) + factor(seller_type)  

  + factor(transmission)+ factor(owner)+ mileage +  

  ↵ engine + max_power + age*km_driven +  

  ↵ age*factor(fuel) + age*factor(seller_type)  

  + age*factor(transmission) + age*factor(owner) +  

  ↵ age*engine + age*max_power +  

  ↵ km_driven*factor(seller_type)  

  + km_driven*factor(transmission) +  

  ↵ km_driven*factor(transmission) +  

  ↵ km_driven*max_power +  

  ↵ factor(fuel)*factor(transmission)

```

```

+ factor(fuel)*factor(owner) + factor(fuel)*mileage +
  ↵  factor(fuel)*engine + factor(fuel)*max_power +
  ↵  factor(seller_type)*mileage
+ factor(seller_type)*engine +
  ↵  factor(seller_type)*max_power +
  ↵  factor(transmission)*factor(owner) +
  ↵  factor(transmission)*engine
+ factor(transmission)*max_power +
  ↵  factor(owner)*mileage + factor(owner)*engine +
  ↵  mileage*engine + mileage*max_power
+ engine*max_power
,data = Car1)

summary(new_Interac_firstordermodel)

```

```

##
## Call:
## lm(formula = selling_price ~ age + km_driven + factor(fuel) +
##     factor(seller_type) + factor(transmission) + factor(owner) +
##     mileage + engine + max_power + age * km_driven + age * factor(fuel) +
##     age * factor(seller_type) + age * factor(transmission) +
##     age * factor(owner) + age * engine + age * max_power + km_driven *
##     factor(seller_type) + km_driven * factor(transmission) +
##     km_driven * factor(transmission) + km_driven * max_power +
##     factor(fuel) * factor(transmission) + factor(fuel) * factor(owner) +
##     factor(fuel) * mileage + factor(fuel) * engine + factor(fuel) *
##     max_power + factor(seller_type) * mileage + factor(seller_type) *
##     engine + factor(seller_type) * max_power + factor(transmission) *
##     factor(owner) + factor(transmission) * engine + factor(transmission) *
##     max_power + factor(owner) * mileage + factor(owner) * engine +
##     mileage * engine + mileage * max_power + engine * max_power,
##     data = Car1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -43969 -1443      12    1411  35167
##
## Coefficients: (3 not defined because of singularities)
##                                     Estimate
## (Intercept)                         -4.241e+03
## age                                 -1.320e+03
## km_driven                           -6.844e-02
## factor(fuel)Petrol                  -1.713e+04
## factor(seller_type)Individual       7.782e+03

```

## factor(seller_type)Trustmark Dealer	-9.078e+02
## factor(transmission)Manual	1.581e+03
## factor(owner)Fourth & Above Owner	1.104e+04
## factor(owner)Second Owner	5.542e+02
## factor(owner)Test Drive Car	8.141e+05
## factor(owner)Third Owner	1.537e+04
## mileage	8.809e+01
## engine	-1.238e+01
## max_power	4.468e+02
## age:km_driven	1.911e-03
## age:factor(fuel)Petrol	5.713e+02
## age:factor(seller_type)Individual	3.845e+02
## age:factor(seller_type)Trustmark Dealer	3.713e+01
## age:factor(transmission)Manual	8.998e+02
## age:factor(owner)Fourth & Above Owner	2.255e+02
## age:factor(owner)Second Owner	6.777e+01
## age:factor(owner)Test Drive Car	NA
## age:factor(owner)Third Owner	-3.830e+01
## age:engine	4.111e-01
## age:max_power	-2.153e+01
## km_driven:factor(seller_type)Individual	2.209e-02
## km_driven:factor(seller_type)Trustmark Dealer	9.789e-02
## km_driven:factor(transmission)Manual	4.478e-02
## km_driven:max_power	-1.810e-04
## factor(fuel)Petrol:factor(transmission)Manual	2.983e+03
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	-3.046e+03
## factor(fuel)Petrol:factor(owner)Second Owner	-5.053e+02
## factor(fuel)Petrol:factor(owner)Test Drive Car	-2.133e+05
## factor(fuel)Petrol:factor(owner)Third Owner	-1.637e+03
## factor(fuel)Petrol:mileage	3.047e+02
## factor(fuel)Petrol:engine	6.746e+00
## factor(fuel)Petrol:max_power	-5.242e+01
## factor(seller_type)Individual:mileage	-2.290e+02
## factor(seller_type)Trustmark Dealer:mileage	-1.389e+02
## factor(seller_type)Individual:engine	2.452e+00
## factor(seller_type)Trustmark Dealer:engine	2.543e+00
## factor(seller_type)Individual:max_power	-1.081e+02
## factor(seller_type)Trustmark Dealer:max_power	-5.107e+01
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	-6.704e+03
## factor(transmission)Manual:factor(owner)Second Owner	-1.149e+03
## factor(transmission)Manual:factor(owner)Test Drive Car	NA
## factor(transmission)Manual:factor(owner)Third Owner	-6.291e+03
## factor(transmission)Manual:engine	1.131e+01
## factor(transmission)Manual:max_power	-2.641e+02
## factor(owner)Fourth & Above Owner:mileage	-1.522e+02

## factor(owner)Second Owner:mileage	2.881e+01
## factor(owner)Test Drive Car:mileage	-3.664e+04
## factor(owner)Third Owner:mileage	-2.737e+02
## factor(owner)Fourth & Above Owner:engine	-2.221e+00
## factor(owner)Second Owner:engine	-3.583e-01
## factor(owner)Test Drive Car:engine	NA
## factor(owner)Third Owner:engine	-2.357e+00
## mileage:engine	-2.865e-01
## mileage:max_power	5.654e+00
## engine:max_power	3.170e-02
##	Std. Error t value
## (Intercept)	2.865e+03 -1.480
## age	1.321e+02 -9.988
## km_driven	8.093e-03 -8.456
## factor(fuel)Petrol	1.533e+03 -11.177
## factor(seller_type)Individual	1.375e+03 5.661
## factor(seller_type)Trustmark Dealer	3.446e+03 -0.263
## factor(transmission)Manual	9.906e+02 1.596
## factor(owner)Fourth & Above Owner	5.011e+03 2.204
## factor(owner)Second Owner	1.788e+03 0.310
## factor(owner)Test Drive Car	8.471e+04 9.611
## factor(owner)Third Owner	3.276e+03 4.693
## mileage	8.692e+01 1.013
## engine	1.390e+00 -8.905
## max_power	1.555e+01 28.740
## age:km_driven	2.591e-04 7.379
## age:factor(fuel)Petrol	4.488e+01 12.729
## age:factor(seller_type)Individual	6.321e+01 6.083
## age:factor(seller_type)Trustmark Dealer	3.499e+02 0.106
## age:factor(transmission)Manual	7.700e+01 11.685
## age:factor(owner)Fourth & Above Owner	1.001e+02 2.252
## age:factor(owner)Second Owner	3.775e+01 1.795
## age:factor(owner)Test Drive Car	NA NA
## age:factor(owner)Third Owner	6.555e+01 -0.584
## age:engine	5.262e-02 7.812
## age:max_power	6.787e-01 -31.728
## km_driven:factor(seller_type)Individual	5.463e-03 4.043
## km_driven:factor(seller_type)Trustmark Dealer	2.985e-02 3.280
## km_driven:factor(transmission)Manual	5.982e-03 7.486
## km_driven:max_power	3.454e-05 -5.240
## factor(fuel)Petrol:factor(transmission)Manual	4.851e+02 6.149
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	1.057e+03 -2.882
## factor(fuel)Petrol:factor(owner)Second Owner	3.567e+02 -1.417
## factor(fuel)Petrol:factor(owner)Test Drive Car	2.582e+04 -8.260
## factor(fuel)Petrol:factor(owner)Third Owner	5.890e+02 -2.780

## factor(fuel)Petrol:mileage	4.376e+01	6.964
## factor(fuel)Petrol:engine	8.084e-01	8.345
## factor(fuel)Petrol:max_power	8.799e+00	-5.958
## factor(seller_type)Individual:mileage	4.365e+01	-5.246
## factor(seller_type)Trustmark Dealer:mileage	1.420e+02	-0.978
## factor(seller_type)Individual:engine	6.046e-01	4.055
## factor(seller_type)Trustmark Dealer:engine	1.364e+00	1.864
## factor(seller_type)Individual:max_power	6.617e+00	-16.343
## factor(seller_type)Trustmark Dealer:max_power	2.627e+01	-1.944
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	1.769e+03	-3.790
## factor(transmission)Manual:factor(owner)Second Owner	4.895e+02	-2.347
## factor(transmission)Manual:factor(owner)Test Drive Car	NA	NA
## factor(transmission)Manual:factor(owner)Third Owner	1.081e+03	-5.819
## factor(transmission)Manual:engine	6.674e-01	16.943
## factor(transmission)Manual:max_power	7.302e+00	-36.172
## factor(owner)Fourth & Above Owner:mileage	1.410e+02	-1.079
## factor(owner)Second Owner:mileage	4.861e+01	0.593
## factor(owner)Test Drive Car:mileage	3.799e+03	-9.645
## factor(owner)Third Owner:mileage	9.360e+01	-2.924
## factor(owner)Fourth & Above Owner:engine	1.152e+00	-1.929
## factor(owner)Second Owner:engine	4.390e-01	-0.816
## factor(owner)Test Drive Car:engine	NA	NA
## factor(owner)Third Owner:engine	7.617e-01	-3.094
## mileage:engine	5.443e-02	-5.264
## mileage:max_power	5.124e-01	11.036
## engine:max_power	3.941e-03	8.045
##	Pr(> t )	
## (Intercept)	0.138841	
## age	< 2e-16	***
## km_driven	< 2e-16	***
## factor(fuel)Petrol	< 2e-16	***
## factor(seller_type)Individual	1.56e-08	***
## factor(seller_type)Trustmark Dealer	0.792250	
## factor(transmission)Manual	0.110547	
## factor(owner)Fourth & Above Owner	0.027571	*
## factor(owner)Second Owner	0.756603	
## factor(owner)Test Drive Car	< 2e-16	***
## factor(owner)Third Owner	2.74e-06	***
## mileage	0.310878	
## engine	< 2e-16	***
## max_power	< 2e-16	***
## age:km_driven	1.77e-13	***
## age:factor(fuel)Petrol	< 2e-16	***
## age:factor(seller_type)Individual	1.23e-09	***
## age:factor(seller_type)Trustmark Dealer	0.915484	

```

## age:factor(transmission)Manual < 2e-16 ***
## age:factor(owner)Fourth & Above Owner 0.024329 *
## age:factor(owner)Second Owner 0.072675 .
## age:factor(owner)Test Drive Car NA
## age:factor(owner)Third Owner 0.559025
## age:engine 6.35e-15 ***
## age:max_power < 2e-16 ***
## km_driven:factor(seller_type)Individual 5.32e-05 ***
## km_driven:factor(seller_type)Trustmark Dealer 0.001043 **
## km_driven:factor(transmission)Manual 7.86e-14 ***
## km_driven:max_power 1.65e-07 ***
## factor(fuel)Petrol:factor(transmission)Manual 8.18e-10 ***
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner 0.003957 **
## factor(fuel)Petrol:factor(owner)Second Owner 0.156653
## factor(fuel)Petrol:factor(owner)Test Drive Car < 2e-16 ***
## factor(fuel)Petrol:factor(owner)Third Owner 0.005451 **
## factor(fuel)Petrol:mileage 3.57e-12 ***
## factor(fuel)Petrol:engine < 2e-16 ***
## factor(fuel)Petrol:max_power 2.67e-09 ***
## factor(seller_type)Individual:mileage 1.59e-07 ***
## factor(seller_type)Trustmark Dealer:mileage 0.328208
## factor(seller_type)Individual:engine 5.07e-05 ***
## factor(seller_type)Trustmark Dealer:engine 0.062291 .
## factor(seller_type)Individual:max_power < 2e-16 ***
## factor(seller_type)Trustmark Dealer:max_power 0.051929 .
## factor(transmission)Manual:factor(owner)Fourth & Above Owner 0.000152 ***
## factor(transmission)Manual:factor(owner)Second Owner 0.018958 *
## factor(transmission)Manual:factor(owner)Test Drive Car NA
## factor(transmission)Manual:factor(owner)Third Owner 6.15e-09 ***
## factor(transmission)Manual:engine < 2e-16 ***
## factor(transmission)Manual:max_power < 2e-16 ***
## factor(owner)Fourth & Above Owner:mileage 0.280508
## factor(owner)Second Owner:mileage 0.553337
## factor(owner)Test Drive Car:mileage < 2e-16 ***
## factor(owner)Third Owner:mileage 0.003463 **
## factor(owner)Fourth & Above Owner:engine 0.053798 .
## factor(owner)Second Owner:engine 0.414421
## factor(owner)Test Drive Car:engine NA
## factor(owner)Third Owner:engine 0.001981 **
## mileage:engine 1.45e-07 ***
## mileage:max_power < 2e-16 ***
## engine:max_power 9.92e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

```

## Residual standard error: 4059 on 7762 degrees of freedom
## Multiple R-squared:  0.9118, Adjusted R-squared:  0.9112
## F-statistic:  1433 on 56 and 7762 DF,  p-value: < 2.2e-16

```

After refitting the model with only significant interaction terms, the  $R^2_{adj}$  value is 0.9112. This indicates that 91.12% of the variation of selling price of a used car can be explained by the interaction model.

## 10. Create a model fit which contains higher degrees.

*#Create a model fit which contains higher degrees.*

```

Pwer_new_Interac_firstordermodel = lm(selling_price ~ age + I(age^2)
                                     + I(age^3)+I(age^4)+ km_driven + factor(fuel) + factor(seller_type)
                                     + factor(transmission)+ factor(owner)+ mileage +
                                     ↵   engine + max_power + age*km_driven +
                                     ↵   age*factor(fuel) + age*factor(seller_type)
                                     + age*factor(transmission) + age*factor(owner) +
                                     ↵   age*engine + age*max_power +
                                     ↵   km_driven*factor(seller_type)
                                     + km_driven*factor(transmission) +
                                     ↵   km_driven*factor(transmission) +
                                     ↵   km_driven*max_power +
                                     ↵   factor(fuel)*factor(transmission)
                                     + factor(fuel)*factor(owner) + factor(fuel)*mileage +
                                     ↵   factor(fuel)*engine + factor(fuel)*max_power +
                                     ↵   factor(seller_type)*mileage
                                     + factor(seller_type)*engine +
                                     ↵   factor(seller_type)*max_power +
                                     ↵   factor(transmission)*factor(owner) +
                                     ↵   factor(transmission)*engine
                                     + factor(transmission)*max_power +
                                     ↵   factor(owner)*mileage + factor(owner)*engine +
                                     ↵   mileage*engine + mileage*max_power
                                     + engine*max_power
                                     ,data = Car1)

summary(Pwer_new_Interac_firstordermodel)

```

##

```

## Call:
## lm(formula = selling_price ~ age + I(age^2) + I(age^3) + I(age^4) +
##     km_driven + factor(fuel) + factor(seller_type) + factor(transmission) +
##     factor(owner) + mileage + engine + max_power + age * km_driven +
##     age * factor(fuel) + age * factor(seller_type) + age * factor(transmission) +
##     age * factor(owner) + age * engine + age * max_power + km_driven *
##     factor(seller_type) + km_driven * factor(transmission) +
##     km_driven * factor(transmission) + km_driven * max_power +
##     factor(fuel) * factor(transmission) + factor(fuel) * factor(owner) +
##     factor(fuel) * mileage + factor(fuel) * engine + factor(fuel) *
##     max_power + factor(seller_type) * mileage + factor(seller_type) *
##     engine + factor(seller_type) * max_power + factor(transmission) *
##     factor(owner) + factor(transmission) * engine + factor(transmission) *
##     max_power + factor(owner) * mileage + factor(owner) * engine +
##     mileage * engine + mileage * max_power + engine * max_power,
##     data = Car1)
##
## Residuals:
##    Min      1Q Median      3Q      Max
## -43299   -1509     -18    1468  35231
##
## Coefficients: (3 not defined because of singularities)
##                                     Estimate
## (Intercept)                   -3.768e+03
## age                      -2.140e+03
## I(age^2)                     1.667e+02
## I(age^3)                     -9.582e+00
## I(age^4)                      1.639e-01
## km_driven                    -6.554e-02
## factor(fuel)Petrol           -1.683e+04
## factor(seller_type)Individual 7.126e+03
## factor(seller_type)Trustmark Dealer -9.286e+02
## factor(transmission)Manual      1.658e+03
## factor(owner)Fourth & Above Owner 9.921e+03
## factor(owner)Second Owner       5.620e+02
## factor(owner)Test Drive Car      8.148e+05
## factor(owner)Third Owner        1.498e+04
## mileage                      1.105e+02
## engine                         -1.199e+01
## max_power                      4.464e+02
## age:km_driven                  1.543e-03
## age:factor(fuel)Petrol          5.924e+02
## age:factor(seller_type)Individual 3.946e+02
## age:factor(seller_type)Trustmark Dealer 3.166e+01
## age:factor(transmission)Manual 8.164e+02

```

## age:factor(owner)Fourth & Above Owner	3.205e+02
## age:factor(owner)Second Owner	5.001e+01
## age:factor(owner)Test Drive Car	NA
## age:factor(owner)Third Owner	-2.245e+01
## age:engine	4.112e-01
## age:max_power	-2.233e+01
## km_driven:factor(seller_type)Individual	2.065e-02
## km_driven:factor(seller_type)Trustmark Dealer	1.033e-01
## km_driven:factor(transmission)Manual	4.572e-02
## km_driven:max_power	-1.698e-04
## factor(fuel)Petrol:factor(transmission)Manual	3.042e+03
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	-3.292e+03
## factor(fuel)Petrol:factor(owner)Second Owner	-5.832e+02
## factor(fuel)Petrol:factor(owner)Test Drive Car	-2.134e+05
## factor(fuel)Petrol:factor(owner)Third Owner	-1.843e+03
## factor(fuel)Petrol:mileage	2.914e+02
## factor(fuel)Petrol:engine	6.628e+00
## factor(fuel)Petrol:max_power	-5.240e+01
## factor(seller_type)Individual:mileage	-2.063e+02
## factor(seller_type)Trustmark Dealer:mileage	-1.609e+02
## factor(seller_type)Individual:engine	2.532e+00
## factor(seller_type)Trustmark Dealer:engine	2.213e+00
## factor(seller_type)Individual:max_power	-1.075e+02
## factor(seller_type)Trustmark Dealer:max_power	-4.306e+01
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	-7.053e+03
## factor(transmission)Manual:factor(owner)Second Owner	-1.111e+03
## factor(transmission)Manual:factor(owner)Test Drive Car	NA
## factor(transmission)Manual:factor(owner)Third Owner	-6.011e+03
## factor(transmission)Manual:engine	1.116e+01
## factor(transmission)Manual:max_power	-2.600e+02
## factor(owner)Fourth & Above Owner:mileage	-1.349e+02
## factor(owner)Second Owner:mileage	3.082e+01
## factor(owner)Test Drive Car:mileage	-3.670e+04
## factor(owner)Third Owner:mileage	-2.779e+02
## factor(owner)Fourth & Above Owner:engine	-2.181e+00
## factor(owner)Second Owner:engine	-3.403e-01
## factor(owner)Test Drive Car:engine	NA
## factor(owner)Third Owner:engine	-2.383e+00
## mileage:engine	-2.984e-01
## mileage:max_power	5.646e+00
## engine:max_power	3.184e-02
##	Std. Error t value
## (Intercept)	2.859e+03 -1.318
## age	2.026e+02 -10.564
## I(age^2)	2.985e+01 5.585

## I(age^3)	2.104e+00	-4.555
## I(age^4)	4.847e-02	3.382
## km_driven	8.186e-03	-8.006
## factor(fuel)Petrol	1.528e+03	-11.017
## factor(seller_type)Individual	1.371e+03	5.197
## factor(seller_type)Trustmark Dealer	3.433e+03	-0.270
## factor(transmission)Manual	9.864e+02	1.681
## factor(owner)Fourth & Above Owner	4.992e+03	1.987
## factor(owner)Second Owner	1.784e+03	0.315
## factor(owner)Test Drive Car	8.434e+04	9.662
## factor(owner)Third Owner	3.266e+03	4.589
## mileage	8.661e+01	1.276
## engine	1.387e+00	-8.643
## max_power	1.549e+01	28.811
## age:km_driven	2.843e-04	5.428
## age:factor(fuel)Petrol	4.617e+01	12.830
## age:factor(seller_type)Individual	6.311e+01	6.252
## age:factor(seller_type)Trustmark Dealer	3.486e+02	0.091
## age:factor(transmission)Manual	7.734e+01	10.556
## age:factor(owner)Fourth & Above Owner	1.040e+02	3.081
## age:factor(owner)Second Owner	3.996e+01	1.251
## age:factor(owner)Test Drive Car		NA NA
## age:factor(owner)Third Owner	6.855e+01	-0.328
## age:engine	5.324e-02	7.723
## age:max_power	7.194e-01	-31.042
## km_driven:factor(seller_type)Individual	5.453e-03	3.787
## km_driven:factor(seller_type)Trustmark Dealer	2.973e-02	3.476
## km_driven:factor(transmission)Manual	5.960e-03	7.671
## km_driven:max_power	3.499e-05	-4.854
## factor(fuel)Petrol:factor(transmission)Manual	4.834e+02	6.293
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	1.054e+03	-3.124
## factor(fuel)Petrol:factor(owner)Second Owner	3.554e+02	-1.641
## factor(fuel)Petrol:factor(owner)Test Drive Car	2.571e+04	-8.300
## factor(fuel)Petrol:factor(owner)Third Owner	5.883e+02	-3.134
## factor(fuel)Petrol:mileage	4.360e+01	6.684
## factor(fuel)Petrol:engine	8.050e-01	8.233
## factor(fuel)Petrol:max_power	8.762e+00	-5.980
## factor(seller_type)Individual:mileage	4.356e+01	-4.737
## factor(seller_type)Trustmark Dealer:mileage	1.415e+02	-1.137
## factor(seller_type)Individual:engine	6.021e-01	4.205
## factor(seller_type)Trustmark Dealer:engine	1.359e+00	1.629
## factor(seller_type)Individual:max_power	6.590e+00	-16.314
## factor(seller_type)Trustmark Dealer:max_power	2.619e+01	-1.644
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	1.771e+03	-3.982
## factor(transmission)Manual:factor(owner)Second Owner	4.881e+02	-2.276

## factor(transmission)Manual:factor(owner)Test Drive Car	NA	NA
## factor(transmission)Manual:factor(owner)Third Owner	1.077e+03	-5.581
## factor(transmission)Manual:engine	6.648e-01	16.784
## factor(transmission)Manual:max_power	7.290e+00	-35.661
## factor(owner)Fourth & Above Owner:mileage	1.405e+02	-0.961
## factor(owner)Second Owner:mileage	4.842e+01	0.637
## factor(owner)Test Drive Car:mileage	3.783e+03	-9.703
## factor(owner)Third Owner:mileage	9.320e+01	-2.982
## factor(owner)Fourth & Above Owner:engine	1.147e+00	-1.901
## factor(owner)Second Owner:engine	4.374e-01	-0.778
## factor(owner)Test Drive Car:engine	NA	NA
## factor(owner)Third Owner:engine	7.586e-01	-3.142
## mileage:engine	5.423e-02	-5.501
## mileage:max_power	5.101e-01	11.068
## engine:max_power	3.927e-03	8.108
##	Pr(> t )	
## (Intercept)	0.187621	
## age	< 2e-16 ***	
## I(age^2)	2.42e-08 ***	
## I(age^3)	5.33e-06 ***	
## I(age^4)	0.000724 ***	
## km_driven	1.36e-15 ***	
## factor(fuel)Petrol	< 2e-16 ***	
## factor(seller_type)Individual	2.08e-07 ***	
## factor(seller_type)Trustmark Dealer	0.786804	
## factor(transmission)Manual	0.092862 .	
## factor(owner)Fourth & Above Owner	0.046907 *	
## factor(owner)Second Owner	0.752702	
## factor(owner)Test Drive Car	< 2e-16 ***	
## factor(owner)Third Owner	4.53e-06 ***	
## mileage	0.201999	
## engine	< 2e-16 ***	
## max_power	< 2e-16 ***	
## age:km_driven	5.87e-08 ***	
## age:factor(fuel)Petrol	< 2e-16 ***	
## age:factor(seller_type)Individual	4.25e-10 ***	
## age:factor(seller_type)Trustmark Dealer	0.927640	
## age:factor(transmission)Manual	< 2e-16 ***	
## age:factor(owner)Fourth & Above Owner	0.002071 **	
## age:factor(owner)Second Owner	0.210806	
## age:factor(owner)Test Drive Car	NA	
## age:factor(owner)Third Owner	0.743252	
## age:engine	1.28e-14 ***	
## age:max_power	< 2e-16 ***	
## km_driven:factor(seller_type)Individual	0.000153 ***	

```

## km_driven:factor(seller_type)Trustmark Dealer          0.000511 ***
## km_driven:factor(transmission)Manual                 1.91e-14 ***
## km_driven:max_power                                1.23e-06 ***
## factor(fuel)Petrol:factor(transmission)Manual       3.29e-10 ***
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner 0.001788 **
## factor(fuel)Petrol:factor(owner)Second Owner        0.100866
## factor(fuel)Petrol:factor(owner)Test Drive Car      < 2e-16 ***
## factor(fuel)Petrol:factor(owner)Third Owner         0.001733 **
## factor(fuel)Petrol:mileage                          2.49e-11 ***
## factor(fuel)Petrol:engine                           < 2e-16 ***
## factor(fuel)Petrol:max_power                      2.32e-09 ***
## factor(seller_type)Individual:mileage              2.21e-06 ***
## factor(seller_type)Trustmark Dealer:mileage        0.255443
## factor(seller_type)Individual:engine                2.64e-05 ***
## factor(seller_type)Trustmark Dealer:engine         0.103365
## factor(seller_type)Individual:max_power            < 2e-16 ***
## factor(seller_type)Trustmark Dealer:max_power      0.100163
## factor(transmission)Manual:factor(owner)Fourth & Above Owner 6.88e-05 ***
## factor(transmission)Manual:factor(owner)Second Owner 0.022903 *
## factor(transmission)Manual:factor(owner)Test Drive Car NA
## factor(transmission)Manual:factor(owner)Third Owner  2.47e-08 ***
## factor(transmission)Manual:engine                  < 2e-16 ***
## factor(transmission)Manual:max_power              < 2e-16 ***
## factor(owner)Fourth & Above Owner:mileage        0.336822
## factor(owner)Second Owner:mileage                 0.524423
## factor(owner)Test Drive Car:mileage               < 2e-16 ***
## factor(owner)Third Owner:mileage                  0.002875 **
## factor(owner)Fourth & Above Owner:engine         0.057379 .
## factor(owner)Second Owner:engine                  0.436542
## factor(owner)Test Drive Car:engine                NA
## factor(owner)Third Owner:engine                  0.001687 **
## mileage:engine                                    3.89e-08 ***
## mileage:max_power                                 < 2e-16 ***
## engine:max_power                                5.95e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4041 on 7759 degrees of freedom
## Multiple R-squared:  0.9126, Adjusted R-squared:  0.9119
## F-statistic: 1373 on 59 and 7759 DF,  p-value: < 2.2e-16

```

Upon adding higher degrees of predictors, we find that there is no significant change in the  $R^2_{adj}$  values. By keeping these terms, we also risk the chance of overfitting the model. Hence, we do not add higher order terms of the predictors to the model.

```
#Create a model fit which contains higher degrees.
```

```
Pwer_new_Interac_firstrdermodel1 = lm(selling_price ~ age + km_driven+
  ↵ I(km_driven^2)+ I(km_driven^3)+ I(km_driven^4) + factor(fuel) +
  ↵ factor(seller_type)
  + factor(transmission)+ factor(owner)+ mileage +
  ↵ engine + max_power + age*km_driven +
  ↵ age*factor(fuel) + age*factor(seller_type)
  + age*factor(transmission) + age*factor(owner) +
  ↵ age*engine + age*max_power +
  ↵ km_driven*factor(seller_type)
  + km_driven*factor(transmission) +
  ↵ km_driven*factor(transmission) +
  ↵ km_driven*max_power +
  ↵ factor(fuel)*factor(transmission)
  + factor(fuel)*factor(owner) + factor(fuel)*mileage +
  ↵ factor(fuel)*engine + factor(fuel)*max_power +
  ↵ factor(seller_type)*mileage
  + factor(seller_type)*engine +
  ↵ factor(seller_type)*max_power +
  ↵ factor(transmission)*factor(owner) +
  ↵ factor(transmission)*engine
  + factor(transmission)*max_power +
  ↵ factor(owner)*mileage + factor(owner)*engine +
  ↵ mileage*engine + mileage*max_power
  + engine*max_power
, data = Car1)
```

```
summary(Pwer_new_Interac_firstrdermodel1)
```

```
##  
## Call:  
## lm(formula = selling_price ~ age + km_driven + I(km_driven^2) +  
##     I(km_driven^3) + I(km_driven^4) + factor(fuel) + factor(seller_type) +  
##     factor(transmission) + factor(owner) + mileage + engine +  
##     max_power + age * km_driven + age * factor(fuel) + age *  
##     factor(seller_type) + age * factor(transmission) + age *  
##     factor(owner) + age * engine + age * max_power + km_driven *  
##     factor(seller_type) + km_driven * factor(transmission) +
```

```

##      km_driven * factor(transmission) + km_driven * max_power +
##      factor(fuel) * factor(transmission) + factor(fuel) * factor(owner) +
##      factor(fuel) * mileage + factor(fuel) * engine + factor(fuel) *
##      max_power + factor(seller_type) * mileage + factor(seller_type) *
##      engine + factor(seller_type) * max_power + factor(transmission) *
##      factor(owner) + factor(transmission) * engine + factor(transmission) *
##      max_power + factor(owner) * mileage + factor(owner) * engine +
##      mileage * engine + mileage * max_power + engine * max_power,
##      data = Car1)
##
## Residuals:
##      Min    1Q Median    3Q   Max
## -44635 -1465     23   1407 35204
##
## Coefficients: (3 not defined because of singularities)
##                                     Estimate
## (Intercept)                      -4.527e+03
## age                         -1.412e+03
## km_driven                     -5.470e-02
## I(km_driven^2)                  9.266e-09
## I(km_driven^3)                  1.225e-14
## I(km_driven^4)                  -7.352e-21
## factor(fuel)Petrol              -1.681e+04
## factor(seller_type)Individual    7.778e+03
## factor(seller_type)Trustmark Dealer -1.137e+03
## factor(transmission)Manual       1.577e+03
## factor(owner)Fourth & Above Owner 1.140e+04
## factor(owner)Second Owner        4.367e+02
## factor(owner)Test Drive Car      8.144e+05
## factor(owner)Third Owner         1.577e+04
## mileage                        9.153e+01
## engine                          -1.222e+01
## max_power                       4.449e+02
## age:km_driven                   2.483e-03
## age:factor(fuel)Petrol           5.802e+02
## age:factor(seller_type)Individual 3.821e+02
## age:factor(seller_type)Trustmark Dealer 8.333e+00
## age:factor(transmission)Manual   9.354e+02
## age:factor(owner)Fourth & Above Owner 1.953e+02
## age:factor(owner)Second Owner    5.167e+01
## age:factor(owner)Test Drive Car  NA
## age:factor(owner)Third Owner     -6.077e+01
## age:engine                       3.717e-01
## age:max_power                    -2.054e+01
## km_driven:factor(seller_type)Individual 2.007e-02

```

	Std. Error	t value
## km_driven:factor(seller_type)Trustmark Dealer	9.951e-02	
## km_driven:factor(transmission)Manual	3.703e-02	
## km_driven:max_power	-3.011e-04	
## factor(fuel)Petrol:factor(transmission)Manual	2.845e+03	
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	-3.111e+03	
## factor(fuel)Petrol:factor(owner)Second Owner	-4.692e+02	
## factor(fuel)Petrol:factor(owner)Test Drive Car	-2.131e+05	
## factor(fuel)Petrol:factor(owner)Third Owner	-1.681e+03	
## factor(fuel)Petrol:mileage	3.012e+02	
## factor(fuel)Petrol:engine	6.693e+00	
## factor(fuel)Petrol:max_power	-5.400e+01	
## factor(seller_type)Individual:mileage	-2.255e+02	
## factor(seller_type)Trustmark Dealer:mileage	-1.291e+02	
## factor(seller_type)Individual:engine	2.511e+00	
## factor(seller_type)Trustmark Dealer:engine	2.545e+00	
## factor(seller_type)Individual:max_power	-1.085e+02	
## factor(seller_type)Trustmark Dealer:max_power	-4.963e+01	
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	-6.527e+03	
## factor(transmission)Manual:factor(owner)Second Owner	-1.110e+03	
## factor(transmission)Manual:factor(owner)Test Drive Car	NA	
## factor(transmission)Manual:factor(owner)Third Owner	-6.272e+03	
## factor(transmission)Manual:engine	1.133e+01	
## factor(transmission)Manual:max_power	-2.612e+02	
## factor(owner)Fourth & Above Owner:mileage	-1.656e+02	
## factor(owner)Second Owner:mileage	3.081e+01	
## factor(owner)Test Drive Car:mileage	-3.667e+04	
## factor(owner)Third Owner:mileage	-2.828e+02	
## factor(owner)Fourth & Above Owner:engine	-2.239e+00	
## factor(owner)Second Owner:engine	-2.602e-01	
## factor(owner)Test Drive Car:engine	NA	
## factor(owner)Third Owner:engine	-2.401e+00	
## mileage:engine	-2.940e-01	
## mileage:max_power	5.728e+00	
## engine:max_power	3.281e-02	
##		
## (Intercept)	2.860e+03	-1.583
## age	1.329e+02	-10.621
## km_driven	8.932e-03	-6.124
## I(km_driven^2)	1.876e-08	0.494
## I(km_driven^3)	1.819e-14	0.673
## I(km_driven^4)	4.831e-21	-1.522
## factor(fuel)Petrol	1.531e+03	-10.977
## factor(seller_type)Individual	1.372e+03	5.670
## factor(seller_type)Trustmark Dealer	3.440e+03	-0.330
## factor(transmission)Manual	9.898e+02	1.593

## factor(owner)Fourth & Above Owner	5.002e+03	2.279
## factor(owner)Second Owner	1.785e+03	0.245
## factor(owner)Test Drive Car	8.453e+04	9.634
## factor(owner)Third Owner	3.270e+03	4.822
## mileage	8.680e+01	1.055
## engine	1.387e+00	-8.807
## max_power	1.552e+01	28.667
## age:km_driven	3.675e-04	6.758
## age:factor(fuel)Petrol	4.495e+01	12.908
## age:factor(seller_type)Individual	6.344e+01	6.023
## age:factor(seller_type)Trustmark Dealer	3.492e+02	0.024
## age:factor(transmission)Manual	7.776e+01	12.029
## age:factor(owner)Fourth & Above Owner	1.003e+02	1.948
## age:factor(owner)Second Owner	3.791e+01	1.363
## age:factor(owner)Test Drive Car		NA NA
## age:factor(owner)Third Owner	6.563e+01	-0.926
## age:engine	5.329e-02	6.976
## age:max_power	6.990e-01	-29.390
## km_driven:factor(seller_type)Individual	5.502e-03	3.649
## km_driven:factor(seller_type)Trustmark Dealer	2.979e-02	3.340
## km_driven:factor(transmission)Manual	6.261e-03	5.915
## km_driven:max_power	4.302e-05	-7.001
## factor(fuel)Petrol:factor(transmission)Manual	4.854e+02	5.861
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	1.055e+03	-2.949
## factor(fuel)Petrol:factor(owner)Second Owner	3.561e+02	-1.318
## factor(fuel)Petrol:factor(owner)Test Drive Car	2.577e+04	-8.269
## factor(fuel)Petrol:factor(owner)Third Owner	5.880e+02	-2.858
## factor(fuel)Petrol:mileage	4.368e+01	6.895
## factor(fuel)Petrol:engine	8.071e-01	8.294
## factor(fuel)Petrol:max_power	8.790e+00	-6.143
## factor(seller_type)Individual:mileage	4.357e+01	-5.177
## factor(seller_type)Trustmark Dealer:mileage	1.418e+02	-0.911
## factor(seller_type)Individual:engine	6.034e-01	4.161
## factor(seller_type)Trustmark Dealer:engine	1.361e+00	1.870
## factor(seller_type)Individual:max_power	6.604e+00	-16.433
## factor(seller_type)Trustmark Dealer:max_power	2.622e+01	-1.893
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	1.766e+03	-3.697
## factor(transmission)Manual:factor(owner)Second Owner	4.885e+02	-2.271
## factor(transmission)Manual:factor(owner)Test Drive Car		NA NA
## factor(transmission)Manual:factor(owner)Third Owner	1.079e+03	-5.813
## factor(transmission)Manual:engine	6.662e-01	17.007
## factor(transmission)Manual:max_power	7.318e+00	-35.699
## factor(owner)Fourth & Above Owner:mileage	1.408e+02	-1.176
## factor(owner)Second Owner:mileage	4.851e+01	0.635
## factor(owner)Test Drive Car:mileage	3.791e+03	-9.671

## factor(owner)Third Owner:mileage	9.343e+01	-3.027
## factor(owner)Fourth & Above Owner:engine	1.150e+00	-1.948
## factor(owner)Second Owner:engine	4.385e-01	-0.593
## factor(owner)Test Drive Car:engine	NA	NA
## factor(owner)Third Owner:engine	7.603e-01	-3.158
## mileage:engine	5.440e-02	-5.404
## mileage:max_power	5.116e-01	11.197
## engine:max_power	3.938e-03	8.332
##	Pr(> t )	
## (Intercept)	0.113562	
## age	< 2e-16 ***	
## km_driven	9.59e-10 ***	
## I(km_driven^2)	0.621304	
## I(km_driven^3)	0.500666	
## I(km_driven^4)	0.128119	
## factor(fuel)Petrol	< 2e-16 ***	
## factor(seller_type)Individual	1.48e-08 ***	
## factor(seller_type)Trustmark Dealer	0.741083	
## factor(transmission)Manual	0.111135	
## factor(owner)Fourth & Above Owner	0.022688 *	
## factor(owner)Second Owner	0.806717	
## factor(owner)Test Drive Car	< 2e-16 ***	
## factor(owner)Third Owner	1.45e-06 ***	
## mileage	0.291656	
## engine	< 2e-16 ***	
## max_power	< 2e-16 ***	
## age:km_driven	1.50e-11 ***	
## age:factor(fuel)Petrol	< 2e-16 ***	
## age:factor(seller_type)Individual	1.79e-09 ***	
## age:factor(seller_type)Trustmark Dealer	0.980961	
## age:factor(transmission)Manual	< 2e-16 ***	
## age:factor(owner)Fourth & Above Owner	0.051397 .	
## age:factor(owner)Second Owner	0.172998	
## age:factor(owner)Test Drive Car	NA	
## age:factor(owner)Third Owner	0.354494	
## age:engine	3.29e-12 ***	
## age:max_power	< 2e-16 ***	
## km_driven:factor(seller_type)Individual	0.000265 ***	
## km_driven:factor(seller_type)Trustmark Dealer	0.000842 ***	
## km_driven:factor(transmission)Manual	3.45e-09 ***	
## km_driven:max_power	2.76e-12 ***	
## factor(fuel)Petrol:factor(transmission)Manual	4.78e-09 ***	
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	0.003197 **	
## factor(fuel)Petrol:factor(owner)Second Owner	0.187586	
## factor(fuel)Petrol:factor(owner)Test Drive Car	< 2e-16 ***	

```

## factor(fuel)Petrol:factor(owner)Third Owner          0.004269 ***
## factor(fuel)Petrol:mileage                         5.79e-12 ***
## factor(fuel)Petrol:engine                          < 2e-16 ***
## factor(fuel)Petrol:max_power                     8.49e-10 ***
## factor(seller_type)Individual:mileage           2.31e-07 ***
## factor(seller_type)Trustmark Dealer:mileage      0.362377
## factor(seller_type)Individual:engine            3.20e-05 ***
## factor(seller_type)Trustmark Dealer:engine       0.061529 .
## factor(seller_type)Individual:max_power         < 2e-16 ***
## factor(seller_type)Trustmark Dealer:max_power    0.058424 .
## factor(transmission)Manual:factor(owner)Fourth & Above Owner 0.000220 ***
## factor(transmission)Manual:factor(owner)Second Owner   0.023157 *
## factor(transmission)Manual:factor(owner)Test Drive Car NA
## factor(transmission)Manual:factor(owner)Third Owner   6.39e-09 ***
## factor(transmission)Manual:engine                < 2e-16 ***
## factor(transmission)Manual:max_power            < 2e-16 ***
## factor(owner)Fourth & Above Owner:mileage        0.239508
## factor(owner)Second Owner:mileage               0.525404
## factor(owner)Test Drive Car:mileage             < 2e-16 ***
## factor(owner)Third Owner:mileage                0.002477 **
## factor(owner)Fourth & Above Owner:engine        0.051456 .
## factor(owner)Second Owner:engine                0.552909
## factor(owner)Test Drive Car:engine              NA
## factor(owner)Third Owner:engine                0.001595 **
## mileage:engine                                6.72e-08 ***
## mileage:max_power                            < 2e-16 ***
## engine:max_power                           < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4050 on 7759 degrees of freedom
## Multiple R-squared:  0.9122, Adjusted R-squared:  0.9115
## F-statistic:  1366 on 59 and 7759 DF,  p-value: < 2.2e-16

```

Upon adding higher degrees of predictors, we find that there is no significant change in the  $R^2_{adj}$  values. By keeping these terms, we also risk the chance of overfitting the model. Hence, we do not add higher order terms of the predictors to the model.

*#Create a model fit which contains higher degrees.*

```
Pwer_new_Interac_firstordermodel2 = lm(selling_price ~ age + km_driven +
← factor(fuel) + factor(seller_type)
```

```

+ factor(transmission)+ factor(owner)+ mileage +
  ↵ I(mileage^2) + I(mileage^3)+ I(mileage^4) +
  ↵ engine + max_power + age*km_driven +
  ↵ age*factor(fuel) + age*factor(seller_type)
+ age*factor(transmission) + age*factor(owner) +
  ↵ age*engine + age*max_power +
  ↵ km_driven*factor(seller_type)
+ km_driven*factor(transmission) +
  ↵ km_driven*factor(transmission) +
  ↵ km_driven*max_power +
  ↵ factor(fuel)*factor(transmission)
+ factor(fuel)*factor(owner) + factor(fuel)*mileage +
  ↵ factor(fuel)*engine + factor(fuel)*max_power +
  ↵ factor(seller_type)*mileage
+ factor(seller_type)*engine +
  ↵ factor(seller_type)*max_power +
  ↵ factor(transmission)*factor(owner) +
  ↵ factor(transmission)*engine
+ factor(transmission)*max_power +
  ↵ factor(owner)*mileage + factor(owner)*engine +
  ↵ mileage*engine + mileage*max_power
+ engine*max_power
,data = Car1)

```

```
summary(Pwer_new_Interac_firstrdermodel2)
```

```

##  

## Call:  

## lm(formula = selling_price ~ age + km_driven + factor(fuel) +  

##      factor(seller_type) + factor(transmission) + factor(owner) +  

##      mileage + I(mileage^2) + I(mileage^3) + I(mileage^4) + engine +  

##      max_power + age * km_driven + age * factor(fuel) + age *  

##      factor(seller_type) + age * factor(transmission) + age *  

##      factor(owner) + age * engine + age * max_power + km_driven *  

##      factor(seller_type) + km_driven * factor(transmission) +  

##      km_driven * factor(transmission) + km_driven * max_power +  

##      factor(fuel) * factor(transmission) + factor(fuel) * factor(owner) +  

##      factor(fuel) * mileage + factor(fuel) * engine + factor(fuel) *  

##      max_power + factor(seller_type) * mileage + factor(seller_type) *  

##      engine + factor(seller_type) * max_power + factor(transmission) *  

##      factor(owner) + factor(transmission) * engine + factor(transmission) *  

##      max_power + factor(owner) * mileage + factor(owner) * engine +

```

```

##      mileage * engine + mileage * max_power + engine * max_power,
##      data = Car1)
##
## Residuals:
##      Min    1Q Median    3Q   Max
## -44045 -1447     -16    1417  35298
##
## Coefficients: (3 not defined because of singularities)
##                                         Estimate
## (Intercept)                      -1.488e+02
## age                         -1.302e+03
## km_driven                     -7.012e-02
## factor(fuel)Petrol             -1.809e+04
## factor(seller_type)Individual  8.535e+03
## factor(seller_type)Trustmark Dealer 1.549e+03
## factor(transmission)Manual    1.528e+03
## factor(owner)Fourth & Above Owner 8.709e+03
## factor(owner)Second Owner     3.766e+02
## factor(owner)Test Drive Car   8.173e+05
## factor(owner)Third Owner      1.448e+04
## mileage                      -1.212e+03
## I(mileage^2)                  1.110e+02
## I(mileage^3)                  -3.936e+00
## I(mileage^4)                  5.045e-02
## engine                        -1.342e+01
## max_power                     4.735e+02
## age:km_driven                1.881e-03
## age:factor(fuel)Petrol        5.804e+02
## age:factor(seller_type)Individual 3.746e+02
## age:factor(seller_type)Trustmark Dealer 6.035e+00
## age:factor(transmission)Manual 8.944e+02
## age:factor(owner)Fourth & Above Owner 2.609e+02
## age:factor(owner)Second Owner  6.975e+01
## age:factor(owner)Test Drive Car NA
## age:factor(owner)Third Owner  -3.240e+01
## age:engine                     4.384e-01
## age:max_power                 -2.193e+01
## km_driven:factor(seller_type)Individual 2.353e-02
## km_driven:factor(seller_type)Trustmark Dealer 1.007e-01
## km_driven:factor(transmission)Manual 4.527e-02
## km_driven:max_power           -1.793e-04
## factor(fuel)Petrol:factor(transmission)Manual 2.929e+03
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner -2.912e+03
## factor(fuel)Petrol:factor(owner)Second Owner    -4.856e+02
## factor(fuel)Petrol:factor(owner)Test Drive Car   -2.140e+05

```

		Std. Error	t value
## factor(fuel)Petrol:factor(owner)Third Owner		-1.563e+03	
## factor(fuel)Petrol:mileage		3.558e+02	
## factor(fuel)Petrol:engine		7.126e+00	
## factor(fuel)Petrol:max_power		-5.662e+01	
## factor(seller_type)Individual:mileage		-2.562e+02	
## factor(seller_type)Trustmark Dealer:mileage		-2.218e+02	
## factor(seller_type)Individual:engine		2.320e+00	
## factor(seller_type)Trustmark Dealer:engine		2.338e+00	
## factor(seller_type)Individual:max_power		-1.084e+02	
## factor(seller_type)Trustmark Dealer:max_power		-5.737e+01	
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	-6.714e+03		
## factor(transmission)Manual:factor(owner)Second Owner	-1.181e+03		
## factor(transmission)Manual:factor(owner)Test Drive Car	NA		
## factor(transmission)Manual:factor(owner)Third Owner	-5.878e+03		
## factor(transmission)Manual:engine	1.118e+01		
## factor(transmission)Manual:max_power	-2.613e+02		
## factor(owner)Fourth & Above Owner:mileage	-7.803e+01		
## factor(owner)Second Owner:mileage	3.452e+01		
## factor(owner)Test Drive Car:mileage	-3.679e+04		
## factor(owner)Third Owner:mileage	-2.550e+02		
## factor(owner)Fourth & Above Owner:engine	-1.817e+00		
## factor(owner)Second Owner:engine	-3.024e-01		
## factor(owner)Test Drive Car:engine	NA		
## factor(owner)Third Owner:engine	-2.315e+00		
## mileage:engine	-1.934e-01		
## mileage:max_power	4.300e+00		
## engine:max_power	2.987e-02		
##			
## (Intercept)	3.572e+03	-0.042	
## age	1.320e+02	-9.862	
## km_driven	8.117e-03	-8.639	
## factor(fuel)Petrol	1.652e+03	-10.950	
## factor(seller_type)Individual	1.428e+03	5.979	
## factor(seller_type)Trustmark Dealer	3.496e+03	0.443	
## factor(transmission)Manual	9.975e+02	1.532	
## factor(owner)Fourth & Above Owner	5.052e+03	1.724	
## factor(owner)Second Owner	1.814e+03	0.208	
## factor(owner)Test Drive Car	8.458e+04	9.663	
## factor(owner)Third Owner	3.290e+03	4.402	
## mileage	2.860e+02	-4.236	
## I(mileage^2)	3.531e+01	3.143	
## I(mileage^3)	1.508e+00	-2.610	
## I(mileage^4)	2.159e-02	2.337	
## engine	1.554e+00	-8.639	
## max_power	1.879e+01	25.199	

## age:km_driven	2.588e-04	7.266
## age:factor(fuel)Petrol	4.496e+01	12.908
## age:factor(seller_type)Individual	6.324e+01	5.923
## age:factor(seller_type)Trustmark Dealer	3.494e+02	0.017
## age:factor(transmission)Manual	7.700e+01	11.616
## age:factor(owner)Fourth & Above Owner	1.002e+02	2.603
## age:factor(owner)Second Owner	3.774e+01	1.848
## age:factor(owner)Test Drive Car	NA	NA
## age:factor(owner)Third Owner	6.547e+01	-0.495
## age:engine	5.320e-02	8.242
## age:max_power	6.865e-01	-31.949
## km_driven:factor(seller_type)Individual	5.474e-03	4.298
## km_driven:factor(seller_type)Trustmark Dealer	2.981e-02	3.378
## km_driven:factor(transmission)Manual	5.975e-03	7.576
## km_driven:max_power	3.464e-05	-5.174
## factor(fuel)Petrol:factor(transmission)Manual	4.922e+02	5.952
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	1.058e+03	-2.752
## factor(fuel)Petrol:factor(owner)Second Owner	3.591e+02	-1.352
## factor(fuel)Petrol:factor(owner)Test Drive Car	2.579e+04	-8.299
## factor(fuel)Petrol:factor(owner)Third Owner	5.903e+02	-2.648
## factor(fuel)Petrol:mileage	5.136e+01	6.928
## factor(fuel)Petrol:engine	8.201e-01	8.689
## factor(fuel)Petrol:max_power	9.108e+00	-6.217
## factor(seller_type)Individual:mileage	4.527e+01	-5.660
## factor(seller_type)Trustmark Dealer:mileage	1.430e+02	-1.551
## factor(seller_type)Individual:engine	6.053e-01	3.833
## factor(seller_type)Trustmark Dealer:engine	1.363e+00	1.716
## factor(seller_type)Individual:max_power	6.765e+00	-16.032
## factor(seller_type)Trustmark Dealer:max_power	2.644e+01	-2.170
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	1.766e+03	-3.801
## factor(transmission)Manual:factor(owner)Second Owner	4.888e+02	-2.416
## factor(transmission)Manual:factor(owner)Test Drive Car	NA	NA
## factor(transmission)Manual:factor(owner)Third Owner	1.085e+03	-5.418
## factor(transmission)Manual:engine	6.864e-01	16.282
## factor(transmission)Manual:max_power	7.458e+00	-35.039
## factor(owner)Fourth & Above Owner:mileage	1.429e+02	-0.546
## factor(owner)Second Owner:mileage	4.966e+01	0.695
## factor(owner)Test Drive Car:mileage	3.794e+03	-9.697
## factor(owner)Third Owner:mileage	9.437e+01	-2.702
## factor(owner)Fourth & Above Owner:engine	1.157e+00	-1.570
## factor(owner)Second Owner:engine	4.433e-01	-0.682
## factor(owner)Test Drive Car:engine	NA	NA
## factor(owner)Third Owner:engine	7.645e-01	-3.028
## mileage:engine	6.733e-02	-2.872
## mileage:max_power	7.107e-01	6.051

## engine:max_power	4.072e-03	7.334
##	Pr(> t )	
## (Intercept)	0.966772	
## age	< 2e-16 ***	
## km_driven	< 2e-16 ***	
## factor(fuel)Petrol	< 2e-16 ***	
## factor(seller_type)Individual	2.35e-09 ***	
## factor(seller_type)Trustmark Dealer	0.657689	
## factor(transmission)Manual	0.125640	
## factor(owner)Fourth & Above Owner	0.084790 .	
## factor(owner)Second Owner	0.835589	
## factor(owner)Test Drive Car	< 2e-16 ***	
## factor(owner)Third Owner	1.09e-05 ***	
## mileage	2.30e-05 ***	
## I(mileage^2)	0.001677 **	
## I(mileage^3)	0.009075 **	
## I(mileage^4)	0.019471 *	
## engine	< 2e-16 ***	
## max_power	< 2e-16 ***	
## age:km_driven	4.06e-13 ***	
## age:factor(fuel)Petrol	< 2e-16 ***	
## age:factor(seller_type)Individual	3.29e-09 ***	
## age:factor(seller_type)Trustmark Dealer	0.986219	
## age:factor(transmission)Manual	< 2e-16 ***	
## age:factor(owner)Fourth & Above Owner	0.009255 **	
## age:factor(owner)Second Owner	0.064592 .	
## age:factor(owner)Test Drive Car	NA	
## age:factor(owner)Third Owner	0.620710	
## age:engine	< 2e-16 ***	
## age:max_power	< 2e-16 ***	
## km_driven:factor(seller_type)Individual	1.75e-05 ***	
## km_driven:factor(seller_type)Trustmark Dealer	0.000734 ***	
## km_driven:factor(transmission)Manual	3.96e-14 ***	
## km_driven:max_power	2.34e-07 ***	
## factor(fuel)Petrol:factor(transmission)Manual	2.77e-09 ***	
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	0.005929 **	
## factor(fuel)Petrol:factor(owner)Second Owner	0.176279	
## factor(fuel)Petrol:factor(owner)Test Drive Car	< 2e-16 ***	
## factor(fuel)Petrol:factor(owner)Third Owner	0.008111 **	
## factor(fuel)Petrol:mileage	4.61e-12 ***	
## factor(fuel)Petrol:engine	< 2e-16 ***	
## factor(fuel)Petrol:max_power	5.34e-10 ***	
## factor(seller_type)Individual:mileage	1.57e-08 ***	
## factor(seller_type)Trustmark Dealer:mileage	0.121024	
## factor(seller_type)Individual:engine	0.000128 ***	

```

## factor(seller_type)Trustmark Dealer:engine          0.086278 .
## factor(seller_type)Individual:max_power           < 2e-16 ***
## factor(seller_type)Trustmark Dealer:max_power      0.030048 *
## factor(transmission)Manual:factor(owner)Fourth & Above Owner 0.000145 ***
## factor(transmission)Manual:factor(owner)Second Owner 0.015734 *
## factor(transmission)Manual:factor(owner)Test Drive Car      NA
## factor(transmission)Manual:factor(owner)Third Owner     6.21e-08 ***
## factor(transmission)Manual:engine                  < 2e-16 ***
## factor(transmission)Manual:max_power              < 2e-16 ***
## factor(owner)Fourth & Above Owner:mileage        0.584944
## factor(owner)Second Owner:mileage                 0.487025
## factor(owner)Test Drive Car:mileage               < 2e-16 ***
## factor(owner)Third Owner:mileage                  0.006911 **
## factor(owner)Fourth & Above Owner:engine         0.116445
## factor(owner)Second Owner:engine                  0.495163
## factor(owner)Test Drive Car:engine                NA
## factor(owner)Third Owner:engine                  0.002467 **
## mileage:engine                                    0.004096 **
## mileage:max_power                                1.51e-09 ***
## engine:max_power                                 2.45e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4053 on 7759 degrees of freedom
## Multiple R-squared:  0.9121, Adjusted R-squared:  0.9114
## F-statistic: 1364 on 59 and 7759 DF,  p-value: < 2.2e-16

```

Upon adding higher degrees of predictors, we find that there is no significant change in the  $R^2_{adj}$  values. By keeping these terms, we also risk the chance of overfitting the model. Hence, we do not add higher order terms of the predictors to the model.

*#Create a model fit which contains higher degrees.*

```

Pwer_new_Interac_firstordermodel3 = lm(selling_price ~ age + km_driven +
  ↵ factor(fuel) + factor(seller_type)
    + factor(transmission)+ factor(owner)+ mileage +
  ↵ engine + I(engine^2)+ I(engine^3) + I(engine^4)
  ↵ + max_power + age*km_driven + age*factor(fuel) +
  ↵ age*factor(seller_type)
  + age*factor(transmission) + age*factor(owner) +
  ↵ age*engine + age*max_power +
  ↵ km_driven*factor(seller_type)

```

```

+ km_driven*factor(transmission) +
  ↵ km_driven*factor(transmission) +
  ↵ km_driven*max_power +
  ↵ factor(fuel)*factor(transmission)
+ factor(fuel)*factor(owner) + factor(fuel)*mileage +
  ↵ factor(fuel)*engine + factor(fuel)*max_power +
  ↵ factor(seller_type)*mileage
+ factor(seller_type)*engine +
  ↵ factor(seller_type)*max_power +
  ↵ factor(transmission)*factor(owner) +
  ↵ factor(transmission)*engine
+ factor(transmission)*max_power +
  ↵ factor(owner)*mileage + factor(owner)*engine +
  ↵ mileage*engine + mileage*max_power
+ engine*max_power
,data = Car1)

```

```
summary(Pwer_new_Interac_firstdordermodel3)
```

```

##
## Call:
## lm(formula = selling_price ~ age + km_driven + factor(fuel) +
##     factor(seller_type) + factor(transmission) + factor(owner) +
##     mileage + engine + I(engine^2) + I(engine^3) + I(engine^4) +
##     max_power + age * km_driven + age * factor(fuel) + age *
##     factor(seller_type) + age * factor(transmission) + age *
##     factor(owner) + age * engine + age * max_power + km_driven *
##     factor(seller_type) + km_driven * factor(transmission) +
##     km_driven * factor(transmission) + km_driven * max_power +
##     factor(fuel) * factor(transmission) + factor(fuel) * factor(owner) +
##     factor(fuel) * mileage + factor(fuel) * engine + factor(fuel) *
##     max_power + factor(seller_type) * mileage + factor(seller_type) *
##     engine + factor(seller_type) * max_power + factor(transmission) *
##     factor(owner) + factor(transmission) * engine + factor(transmission) *
##     max_power + factor(owner) * mileage + factor(owner) * engine +
##     mileage * engine + mileage * max_power + engine * max_power,
##     data = Car1)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -44270   -1422      32    1389   34896
##
```

```

## Coefficients: (3 not defined because of singularities)
##                                     Estimate
## (Intercept)                         -1.148e+04
## age                                -1.218e+03
## km_driven                           -6.758e-02
## factor(fuel)Petrol                  -2.235e+04
## factor(seller_type)Individual       5.523e+03
## factor(seller_type)Trustmark Dealer 3.481e+03
## factor(transmission)Manual          1.669e+01
## factor(owner)Fourth & Above Owner 1.092e+04
## factor(owner)Second Owner          3.991e+02
## factor(owner)Test Drive Car        8.150e+05
## factor(owner)Third Owner           1.514e+04
## mileage                             -1.001e+02
## engine                              2.899e+01
## I(engine^2)                          -4.926e-02
## I(engine^3)                          2.035e-05
## I(engine^4)                          -2.815e-09
## max_power                           4.855e+02
## age:km_driven                      1.906e-03
## age:factor(fuel)Petrol              5.452e+02
## age:factor(seller_type)Individual  3.977e+02
## age:factor(seller_type)Trustmark Dealer 7.103e+01
## age:factor(transmission)Manual     8.405e+02
## age:factor(owner)Fourth & Above Owner 2.177e+02
## age:factor(owner)Second Owner      6.833e+01
## age:factor(owner)Test Drive Car    NA
## age:factor(owner)Third Owner       -4.272e+01
## age:engine                           4.211e-01
## age:max_power                       -2.212e+01
## km_driven:factor(seller_type)Individual 1.923e-02
## km_driven:factor(seller_type)Trustmark Dealer 8.521e-02
## km_driven:factor(transmission)Manual 4.872e-02
## km_driven:max_power                -2.043e-04
## factor(fuel)Petrol:factor(transmission)Manual 3.697e+03
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner -2.909e+03
## factor(fuel)Petrol:factor(owner)Second Owner   -5.208e+02
## factor(fuel)Petrol:factor(owner)Test Drive Car -2.133e+05
## factor(fuel)Petrol:factor(owner)Third Owner    -1.437e+03
## factor(fuel)Petrol:mileage          3.440e+02
## factor(fuel)Petrol:engine          1.042e+01
## factor(fuel)Petrol:max_power      -6.802e+01
## factor(seller_type)Individual:mileage -1.777e+02
## factor(seller_type)Trustmark Dealer:mileage -2.585e+02
## factor(seller_type)Individual:engine 2.901e+00

```

## factor(seller_type)Trustmark Dealer:engine	2.575e+00
## factor(seller_type)Individual:max_power	-1.012e+02
## factor(seller_type)Trustmark Dealer:max_power	-7.415e+01
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	-6.343e+03
## factor(transmission)Manual:factor(owner)Second Owner	-1.015e+03
## factor(transmission)Manual:factor(owner)Test Drive Car	NA
## factor(transmission)Manual:factor(owner)Third Owner	-6.406e+03
## factor(transmission)Manual:engine	1.204e+01
## factor(transmission)Manual:max_power	-2.630e+02
## factor(owner)Fourth & Above Owner:mileage	-1.671e+02
## factor(owner)Second Owner:mileage	3.224e+01
## factor(owner)Test Drive Car:mileage	-3.666e+04
## factor(owner)Third Owner:mileage	-2.744e+02
## factor(owner)Fourth & Above Owner:engine	-2.212e+00
## factor(owner)Second Owner:engine	-3.853e-01
## factor(owner)Test Drive Car:engine	NA
## factor(owner)Third Owner:engine	-2.127e+00
## mileage:engine	-1.531e-01
## mileage:max_power	4.743e+00
## engine:max_power	2.615e-02
##	Std. Error t value
## (Intercept)	4.292e+03 -2.676
## age	1.321e+02 -9.219
## km_driven	8.087e-03 -8.357
## factor(fuel)Petrol	1.702e+03 -13.133
## factor(seller_type)Individual	1.386e+03 3.986
## factor(seller_type)Trustmark Dealer	3.490e+03 0.998
## factor(transmission)Manual	9.950e+02 0.017
## factor(owner)Fourth & Above Owner	4.977e+03 2.195
## factor(owner)Second Owner	1.776e+03 0.225
## factor(owner)Test Drive Car	8.409e+04 9.692
## factor(owner)Third Owner	3.268e+03 4.633
## mileage	9.598e+01 -1.043
## engine	7.176e+00 4.039
## I(engine^2)	6.045e-03 -8.149
## I(engine^3)	2.215e-06 9.188
## I(engine^4)	2.909e-10 -9.677
## max_power	1.931e+01 25.140
## age:km_driven	2.572e-04 7.410
## age:factor(fuel)Petrol	4.492e+01 12.137
## age:factor(seller_type)Individual	6.279e+01 6.334
## age:factor(seller_type)Trustmark Dealer	3.474e+02 0.204
## age:factor(transmission)Manual	7.683e+01 10.940
## age:factor(owner)Fourth & Above Owner	9.943e+01 2.189
## age:factor(owner)Second Owner	3.748e+01 1.823

## age:factor(owner)Test Drive Car	NA	NA
## age:factor(owner)Third Owner	6.509e+01	-0.656
## age:engine	5.250e-02	8.021
## age:max_power	6.790e-01	-32.586
## km_driven:factor(seller_type)Individual	5.450e-03	3.528
## km_driven:factor(seller_type)Trustmark Dealer	2.965e-02	2.874
## km_driven:factor(transmission)Manual	5.948e-03	8.191
## km_driven:max_power	3.465e-05	-5.897
## factor(fuel)Petrol:factor(transmission)Manual	4.872e+02	7.590
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	1.049e+03	-2.772
## factor(fuel)Petrol:factor(owner)Second Owner	3.541e+02	-1.471
## factor(fuel)Petrol:factor(owner)Test Drive Car	2.564e+04	-8.319
## factor(fuel)Petrol:factor(owner)Third Owner	5.864e+02	-2.450
## factor(fuel)Petrol:mileage	4.578e+01	7.514
## factor(fuel)Petrol:engine	9.599e-01	10.859
## factor(fuel)Petrol:max_power	9.206e+00	-7.389
## factor(seller_type)Individual:mileage	4.372e+01	-4.066
## factor(seller_type)Trustmark Dealer:mileage	1.417e+02	-1.824
## factor(seller_type)Individual:engine	6.025e-01	4.815
## factor(seller_type)Trustmark Dealer:engine	1.354e+00	1.903
## factor(seller_type)Individual:max_power	6.641e+00	-15.234
## factor(seller_type)Trustmark Dealer:max_power	2.645e+01	-2.803
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	1.757e+03	-3.610
## factor(transmission)Manual:factor(owner)Second Owner	4.864e+02	-2.088
## factor(transmission)Manual:factor(owner)Test Drive Car	NA	NA
## factor(transmission)Manual:factor(owner)Third Owner	1.074e+03	-5.964
## factor(transmission)Manual:engine	6.764e-01	17.802
## factor(transmission)Manual:max_power	7.324e+00	-35.903
## factor(owner)Fourth & Above Owner:mileage	1.401e+02	-1.192
## factor(owner)Second Owner:mileage	4.829e+01	0.668
## factor(owner)Test Drive Car:mileage	3.772e+03	-9.719
## factor(owner)Third Owner:mileage	9.343e+01	-2.937
## factor(owner)Fourth & Above Owner:engine	1.144e+00	-1.933
## factor(owner)Second Owner:engine	4.358e-01	-0.884
## factor(owner)Test Drive Car:engine	NA	NA
## factor(owner)Third Owner:engine	7.593e-01	-2.801
## mileage:engine	6.260e-02	-2.446
## mileage:max_power	5.271e-01	8.998
## engine:max_power	5.766e-03	4.534
##	Pr(> t )	
## (Intercept)	0.007475	**
## age	< 2e-16	***
## km_driven	< 2e-16	***
## factor(fuel)Petrol	< 2e-16	***
## factor(seller_type)Individual	6.79e-05	***

## factor(seller_type)Trustmark Dealer	0.318549
## factor(transmission)Manual	0.986616
## factor(owner)Fourth & Above Owner	0.028179 *
## factor(owner)Second Owner	0.822208
## factor(owner)Test Drive Car	< 2e-16 ***
## factor(owner)Third Owner	3.66e-06 ***
## mileage	0.296916
## engine	5.42e-05 ***
## I(engine^2)	4.24e-16 ***
## I(engine^3)	< 2e-16 ***
## I(engine^4)	< 2e-16 ***
## max_power	< 2e-16 ***
## age:km_driven	1.40e-13 ***
## age:factor(fuel)Petrol	< 2e-16 ***
## age:factor(seller_type)Individual	2.53e-10 ***
## age:factor(seller_type)Trustmark Dealer	0.838017
## age:factor(transmission)Manual	< 2e-16 ***
## age:factor(owner)Fourth & Above Owner	0.028593 *
## age:factor(owner)Second Owner	0.068295 .
## age:factor(owner)Test Drive Car	NA
## age:factor(owner)Third Owner	0.511593
## age:engine	1.21e-15 ***
## age:max_power	< 2e-16 ***
## km_driven:factor(seller_type)Individual	0.000421 ***
## km_driven:factor(seller_type)Trustmark Dealer	0.004069 **
## km_driven:factor(transmission)Manual	3.01e-16 ***
## km_driven:max_power	3.87e-09 ***
## factor(fuel)Petrol:factor(transmission)Manual	3.58e-14 ***
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	0.005587 **
## factor(fuel)Petrol:factor(owner)Second Owner	0.141404
## factor(fuel)Petrol:factor(owner)Test Drive Car	< 2e-16 ***
## factor(fuel)Petrol:factor(owner)Third Owner	0.014304 *
## factor(fuel)Petrol:mileage	6.39e-14 ***
## factor(fuel)Petrol:engine	< 2e-16 ***
## factor(fuel)Petrol:max_power	1.64e-13 ***
## factor(seller_type)Individual:mileage	4.84e-05 ***
## factor(seller_type)Trustmark Dealer:mileage	0.068166 .
## factor(seller_type)Individual:engine	1.50e-06 ***
## factor(seller_type)Trustmark Dealer:engine	0.057127 .
## factor(seller_type)Individual:max_power	< 2e-16 ***
## factor(seller_type)Trustmark Dealer:max_power	0.005077 **
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	0.000309 ***
## factor(transmission)Manual:factor(owner)Second Owner	0.036853 *
## factor(transmission)Manual:factor(owner)Test Drive Car	NA
## factor(transmission)Manual:factor(owner)Third Owner	2.57e-09 ***

```

## factor(transmission)Manual:engine < 2e-16 ***
## factor(transmission)Manual:max_power < 2e-16 ***
## factor(owner)Fourth & Above Owner:mileage 0.233281
## factor(owner)Second Owner:mileage 0.504354
## factor(owner)Test Drive Car:mileage < 2e-16 ***
## factor(owner)Third Owner:mileage 0.003325 **
## factor(owner)Fourth & Above Owner:engine 0.053257 .
## factor(owner)Second Owner:engine 0.376731
## factor(owner)Test Drive Car:engine NA
## factor(owner)Third Owner:engine 0.005109 **
## mileage:engine 0.014473 *
## mileage:max_power < 2e-16 ***
## engine:max_power 5.87e-06 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4029 on 7759 degrees of freedom
## Multiple R-squared: 0.9131, Adjusted R-squared: 0.9125
## F-statistic: 1382 on 59 and 7759 DF, p-value: < 2.2e-16

```

Upon adding higher degrees of predictors, we find that there is no significant change in the  $R^2_{adj}$  values. By keeping these terms, we also risk the chance of overfitting the model. Hence, we do not add higher order terms of the predictors to the model.

*#Create a model fit which contains higher degrees.*

```

Pwer_new_Interac_firstordermodel4 = lm(selling_price ~ age + km_driven +
  ↵ factor(fuel) + factor(seller_type)
  + factor(transmission)+ factor(owner)+ mileage +
  ↵ engine + max_power + I(max_power^2) +
  ↵ age*km_driven + age*factor(fuel) +
  ↵ age*factor(seller_type)
  + age*factor(transmission) + age*factor(owner) +
  ↵ age*engine + age*max_power +
  ↵ km_driven*factor(seller_type)
  + km_driven*factor(transmission) +
  ↵ km_driven*factor(transmission) +
  ↵ km_driven*max_power +
  ↵ factor(fuel)*factor(transmission)
  + factor(fuel)*factor(owner) + factor(fuel)*mileage +
  ↵ factor(fuel)*engine + factor(fuel)*max_power +
  ↵ factor(seller_type)*mileage

```

```

+ factor(seller_type)*engine +
  ↳ factor(seller_type)*max_power +
  ↳ factor(transmission)*factor(owner) +
  ↳ factor(transmission)*engine
+ factor(transmission)*max_power +
  ↳ factor(owner)*mileage + factor(owner)*engine +
  ↳ mileage*engine + mileage*max_power
+ engine*max_power
,data = Car1)

```

```
summary(Pwer_new_Interac_firstrdermodel4)
```

```

##
## Call:
## lm(formula = selling_price ~ age + km_driven + factor(fuel) +
##     factor(seller_type) + factor(transmission) + factor(owner) +
##     mileage + engine + max_power + I(max_power^2) + age * km_driven +
##     age * factor(fuel) + age * factor(seller_type) + age * factor(transmission) +
##     age * factor(owner) + age * engine + age * max_power + km_driven *
##     factor(seller_type) + km_driven * factor(transmission) +
##     km_driven * factor(transmission) + km_driven * max_power +
##     factor(fuel) * factor(transmission) + factor(fuel) * factor(owner) +
##     factor(fuel) * mileage + factor(fuel) * engine + factor(fuel) *
##     max_power + factor(seller_type) * mileage + factor(seller_type) *
##     engine + factor(seller_type) * max_power + factor(transmission) *
##     factor(owner) + factor(transmission) * engine + factor(transmission) *
##     max_power + factor(owner) * mileage + factor(owner) * engine +
##     mileage * engine + mileage * max_power + engine * max_power,
##     data = Car1)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -43924   -1443     13   1410  35164
##
## Coefficients: (3 not defined because of singularities)
##                                     Estimate
## (Intercept)                      -4.147e+03
## age                               -1.319e+03
## km_driven                         -6.838e-02
## factor(fuel)Petrol                  -1.715e+04
## factor(seller_type)Individual       7.743e+03
## factor(seller_type)Trustmark Dealer -9.414e+02

```

## factor(transmission)Manual	1.592e+03
## factor(owner)Fourth & Above Owner	1.106e+04
## factor(owner)Second Owner	5.514e+02
## factor(owner)Test Drive Car	8.142e+05
## factor(owner)Third Owner	1.538e+04
## mileage	8.474e+01
## engine	-1.250e+01
## max_power	4.477e+02
## I(max_power^2)	-9.185e-03
## age:km_driven	1.911e-03
## age:factor(fuel)Petrol	5.716e+02
## age:factor(seller_type)Individual	3.850e+02
## age:factor(seller_type)Trustmark Dealer	3.685e+01
## age:factor(transmission)Manual	8.989e+02
## age:factor(owner)Fourth & Above Owner	2.253e+02
## age:factor(owner)Second Owner	6.794e+01
## age:factor(owner)Test Drive Car	NA
## age:factor(owner)Third Owner	-3.830e+01
## age:engine	4.112e-01
## age:max_power	-2.154e+01
## km_driven:factor(seller_type)Individual	2.204e-02
## km_driven:factor(seller_type)Trustmark Dealer	9.782e-02
## km_driven:factor(transmission)Manual	4.480e-02
## km_driven:max_power	-1.813e-04
## factor(fuel)Petrol:factor(transmission)Manual	2.989e+03
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	-3.046e+03
## factor(fuel)Petrol:factor(owner)Second Owner	-5.047e+02
## factor(fuel)Petrol:factor(owner)Test Drive Car	-2.133e+05
## factor(fuel)Petrol:factor(owner)Third Owner	-1.637e+03
## factor(fuel)Petrol:mileage	3.050e+02
## factor(fuel)Petrol:engine	6.735e+00
## factor(fuel)Petrol:max_power	-5.218e+01
## factor(seller_type)Individual:mileage	-2.278e+02
## factor(seller_type)Trustmark Dealer:mileage	-1.373e+02
## factor(seller_type)Individual:engine	2.459e+00
## factor(seller_type)Trustmark Dealer:engine	2.547e+00
## factor(seller_type)Individual:max_power	-1.081e+02
## factor(seller_type)Trustmark Dealer:max_power	-5.106e+01
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	-6.709e+03
## factor(transmission)Manual:factor(owner)Second Owner	-1.149e+03
## factor(transmission)Manual:factor(owner)Test Drive Car	NA
## factor(transmission)Manual:factor(owner)Third Owner	-6.293e+03
## factor(transmission)Manual:engine	1.137e+01
## factor(transmission)Manual:max_power	-2.652e+02
## factor(owner)Fourth & Above Owner:mileage	-1.522e+02

## factor(owner)Second Owner:mileage	2.894e+01
## factor(owner)Test Drive Car:mileage	-3.665e+04
## factor(owner)Third Owner:mileage	-2.738e+02
## factor(owner)Fourth & Above Owner:engine	-2.226e+00
## factor(owner)Second Owner:engine	-3.589e-01
## factor(owner)Test Drive Car:engine	NA
## factor(owner)Third Owner:engine	-2.357e+00
## mileage:engine	-2.871e-01
## mileage:max_power	5.690e+00
## engine:max_power	3.246e-02
##	Std. Error t value
## (Intercept)	2.939e+03 -1.411
## age	1.322e+02 -9.982
## km_driven	8.104e-03 -8.438
## factor(fuel)Petrol	1.539e+03 -11.146
## factor(seller_type)Individual	1.402e+03 5.523
## factor(seller_type)Trustmark Dealer	3.455e+03 -0.273
## factor(transmission)Manual	9.936e+02 1.602
## factor(owner)Fourth & Above Owner	5.012e+03 2.206
## factor(owner)Second Owner	1.788e+03 0.308
## factor(owner)Test Drive Car	8.471e+04 9.611
## factor(owner)Third Owner	3.276e+03 4.694
## mileage	9.007e+01 0.941
## engine	1.643e+00 -7.607
## max_power	1.659e+01 26.989
## I(max_power^2)	6.476e-02 -0.142
## age:km_driven	2.592e-04 7.372
## age:factor(fuel)Petrol	4.491e+01 12.726
## age:factor(seller_type)Individual	6.329e+01 6.082
## age:factor(seller_type)Trustmark Dealer	3.499e+02 0.105
## age:factor(transmission)Manual	7.729e+01 11.630
## age:factor(owner)Fourth & Above Owner	1.001e+02 2.250
## age:factor(owner)Second Owner	3.777e+01 1.799
## age:factor(owner)Test Drive Car	NA NA
## age:factor(owner)Third Owner	6.555e+01 -0.584
## age:engine	5.263e-02 7.813
## age:max_power	6.789e-01 -31.723
## km_driven:factor(seller_type)Individual	5.474e-03 4.027
## km_driven:factor(seller_type)Trustmark Dealer	2.985e-02 3.277
## km_driven:factor(transmission)Manual	5.984e-03 7.487
## km_driven:max_power	3.461e-05 -5.239
## factor(fuel)Petrol:factor(transmission)Manual	4.868e+02 6.139
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	1.057e+03 -2.882
## factor(fuel)Petrol:factor(owner)Second Owner	3.568e+02 -1.415
## factor(fuel)Petrol:factor(owner)Test Drive Car	2.583e+04 -8.260

## factor(fuel)Petrol:factor(owner)Third Owner	5.891e+02	-2.779
## factor(fuel)Petrol:mileage	4.379e+01	6.964
## factor(fuel)Petrol:engine	8.123e-01	8.291
## factor(fuel)Petrol:max_power	8.969e+00	-5.817
## factor(seller_type)Individual:mileage	4.442e+01	-5.129
## factor(seller_type)Trustmark Dealer:mileage	1.424e+02	-0.964
## factor(seller_type)Individual:engine	6.066e-01	4.053
## factor(seller_type)Trustmark Dealer:engine	1.364e+00	1.867
## factor(seller_type)Individual:max_power	6.635e+00	-16.288
## factor(seller_type)Trustmark Dealer:max_power	2.627e+01	-1.944
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	1.769e+03	-3.792
## factor(transmission)Manual:factor(owner)Second Owner	4.895e+02	-2.347
## factor(transmission)Manual:factor(owner)Test Drive Car	NA	NA
## factor(transmission)Manual:factor(owner)Third Owner	1.081e+03	-5.820
## factor(transmission)Manual:engine	8.060e-01	14.108
## factor(transmission)Manual:max_power	1.035e+01	-25.610
## factor(owner)Fourth & Above Owner:mileage	1.410e+02	-1.080
## factor(owner)Second Owner:mileage	4.862e+01	0.595
## factor(owner)Test Drive Car:mileage	3.800e+03	-9.645
## factor(owner)Third Owner:mileage	9.361e+01	-2.925
## factor(owner)Fourth & Above Owner:engine	1.152e+00	-1.932
## factor(owner)Second Owner:engine	4.391e-01	-0.817
## factor(owner)Test Drive Car:engine	NA	NA
## factor(owner)Third Owner:engine	7.618e-01	-3.094
## mileage:engine	5.458e-02	-5.260
## mileage:max_power	5.693e-01	9.995
## engine:max_power	6.670e-03	4.867
##	Pr(> t )	
## (Intercept)	0.158233	
## age	< 2e-16 ***	
## km_driven	< 2e-16 ***	
## factor(fuel)Petrol	< 2e-16 ***	
## factor(seller_type)Individual	3.44e-08 ***	
## factor(seller_type)Trustmark Dealer	0.785238	
## factor(transmission)Manual	0.109198	
## factor(owner)Fourth & Above Owner	0.027411 *	
## factor(owner)Second Owner	0.757846	
## factor(owner)Test Drive Car	< 2e-16 ***	
## factor(owner)Third Owner	2.73e-06 ***	
## mileage	0.346826	
## engine	3.15e-14 ***	
## max_power	< 2e-16 ***	
## I(max_power^2)	0.887208	
## age:km_driven	1.86e-13 ***	
## age:factor(fuel)Petrol	< 2e-16 ***	

```

## age:factor(seller_type)Individual           1.24e-09 ***
## age:factor(seller_type)Trustmark Dealer    0.916140
## age:factor(transmission)Manual            < 2e-16 ***
## age:factor(owner)Fourth & Above Owner    0.024480 *
## age:factor(owner)Second Owner             0.072115 .
## age:factor(owner)Test Drive Car          NA
## age:factor(owner)Third Owner             0.559097
## age:engine                            6.32e-15 ***
## age:max_power                         < 2e-16 ***
## km_driven:factor(seller_type)Individual 5.71e-05 ***
## km_driven:factor(seller_type)Trustmark Dealer 0.001055 **
## km_driven:factor(transmission)Manual     7.81e-14 ***
## km_driven:max_power                   1.66e-07 ***
## factor(fuel)Petrol:factor(transmission)Manual 8.69e-10 ***
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner 0.003962 **
## factor(fuel)Petrol:factor(owner)Second Owner 0.157180
## factor(fuel)Petrol:factor(owner)Test Drive Car < 2e-16 ***
## factor(fuel)Petrol:factor(owner)Third Owner 0.005460 **
## factor(fuel)Petrol:mileage              3.57e-12 ***
## factor(fuel)Petrol:engine               < 2e-16 ***
## factor(fuel)Petrol:max_power          6.21e-09 ***
## factor(seller_type)Individual:mileage   2.98e-07 ***
## factor(seller_type)Trustmark Dealer:mileage 0.334924
## factor(seller_type)Individual:engine    5.11e-05 ***
## factor(seller_type)Trustmark Dealer:engine 0.061938 .
## factor(seller_type)Individual:max_power < 2e-16 ***
## factor(seller_type)Trustmark Dealer:max_power 0.051971 .
## factor(transmission)Manual:factor(owner)Fourth & Above Owner 0.000151 ***
## factor(transmission)Manual:factor(owner)Second Owner 0.018925 *
## factor(transmission)Manual:factor(owner)Test Drive Car NA
## factor(transmission)Manual:factor(owner)Third Owner 6.11e-09 ***
## factor(transmission)Manual:engine        < 2e-16 ***
## factor(transmission)Manual:max_power    < 2e-16 ***
## factor(owner)Fourth & Above Owner:mileage 0.280388
## factor(owner)Second Owner:mileage       0.551680
## factor(owner)Test Drive Car:mileage     < 2e-16 ***
## factor(owner)Third Owner:mileage        0.003457 **
## factor(owner)Fourth & Above Owner:engine 0.053384 .
## factor(owner)Second Owner:engine        0.413754
## factor(owner)Test Drive Car:engine     NA
## factor(owner)Third Owner:engine        0.001985 **
## mileage:engine                        1.48e-07 ***
## mileage:max_power                   < 2e-16 ***
## engine:max_power                  1.15e-06 ***
## ---
```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4059 on 7761 degrees of freedom
## Multiple R-squared:  0.9118, Adjusted R-squared:  0.9111
## F-statistic:  1407 on 57 and 7761 DF,  p-value: < 2.2e-16

```

Upon adding higher degrees of predictors, we find that there is no significant change in the  $R^2_{adj}$  values. By keeping these terms, we also risk the chance of overfitting the model. Hence, we do not add higher order terms of the predictors to the model.

### **11. Compare the $R^2$ adjusted value and RMSE value of the addictive model, interaction model and higher model.**

After fitting the addictive model, the interaction model and higher model, the  $R^2_{adj}$  values are: 0.683, 0.9114 and 0.9111 respectively. This indicates that the best model for predicting the selling price of a used car is the interaction model, since its  $R^2_{adj}$  value of 0.9114 explains 91.14% of the variation of the selling price of a used car. Also the RMSE of the interaction term is 4059, which is lower than that of the addictive model which is 7668.

Finally, after comparing the  $R^2_{adj}$  values and the RMSE of the 3 models, we conclude that the interaction model is the best for predicting the selling price of a used car.

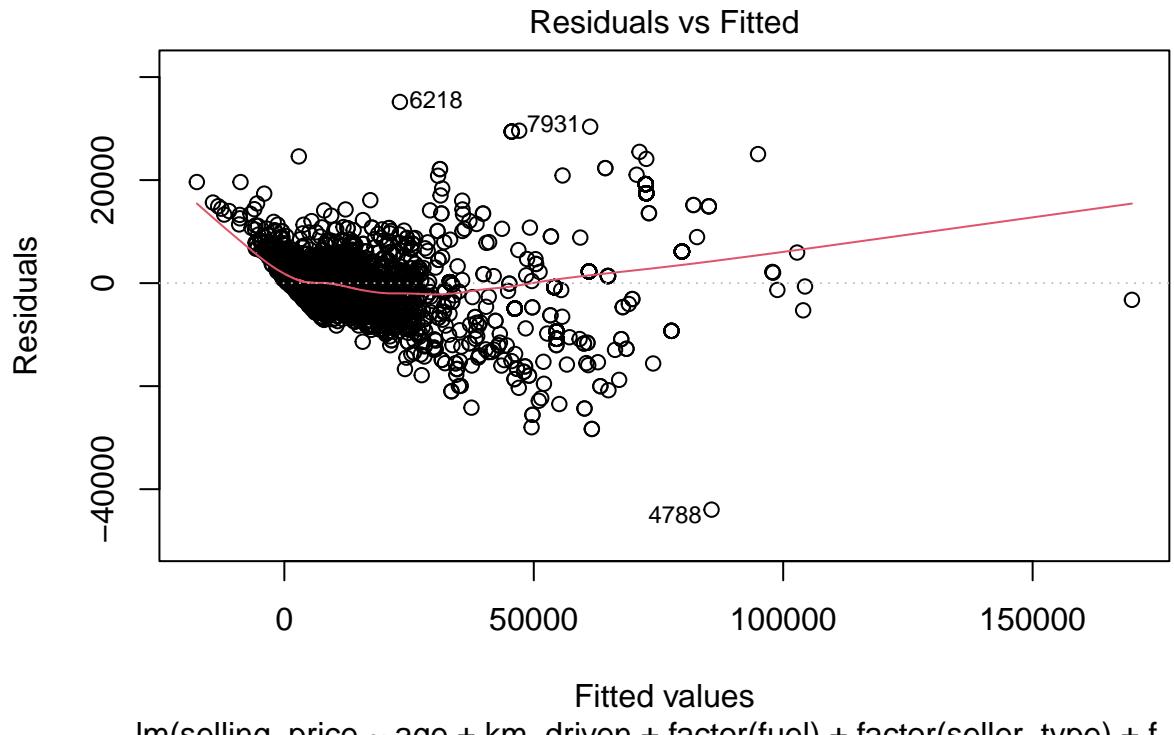
### **12. Test the best model for model assumptions such as equal variance, linearity, normality.**

#### Checking for linearity

```

#Test the best model for model assumptions such as equal variance,
#  linearity, normality,
#
# Linearity
par(mfrow=c(1,1))
plot(new_Interac_firstordermodel, which = 1)

```



From the plot of the residuals vs fitted graph, the error terms appear to form a non linear pattern, indicating the model is not linear.

#### Checking the equal variance assumption

$H_0$ : heteroscedasticity is not present(homoscedasticity)

$H_{alpha}$ : heteroscedasticity is present

```
#Heteroscedasticity
```

```
library(lmtest)
```

```
## Loading required package: zoo

## 
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
## 
##     as.Date, as.Date.numeric
```

```

bptest(new_Interac_firstordermodel)

## 
## studentized Breusch-Pagan test
##
## data: new_Interac_firstordermodel
## BP = 2877.8, df = 56, p-value < 2.2e-16

library(olsrr)
ols_test_f(new_Interac_firstordermodel)

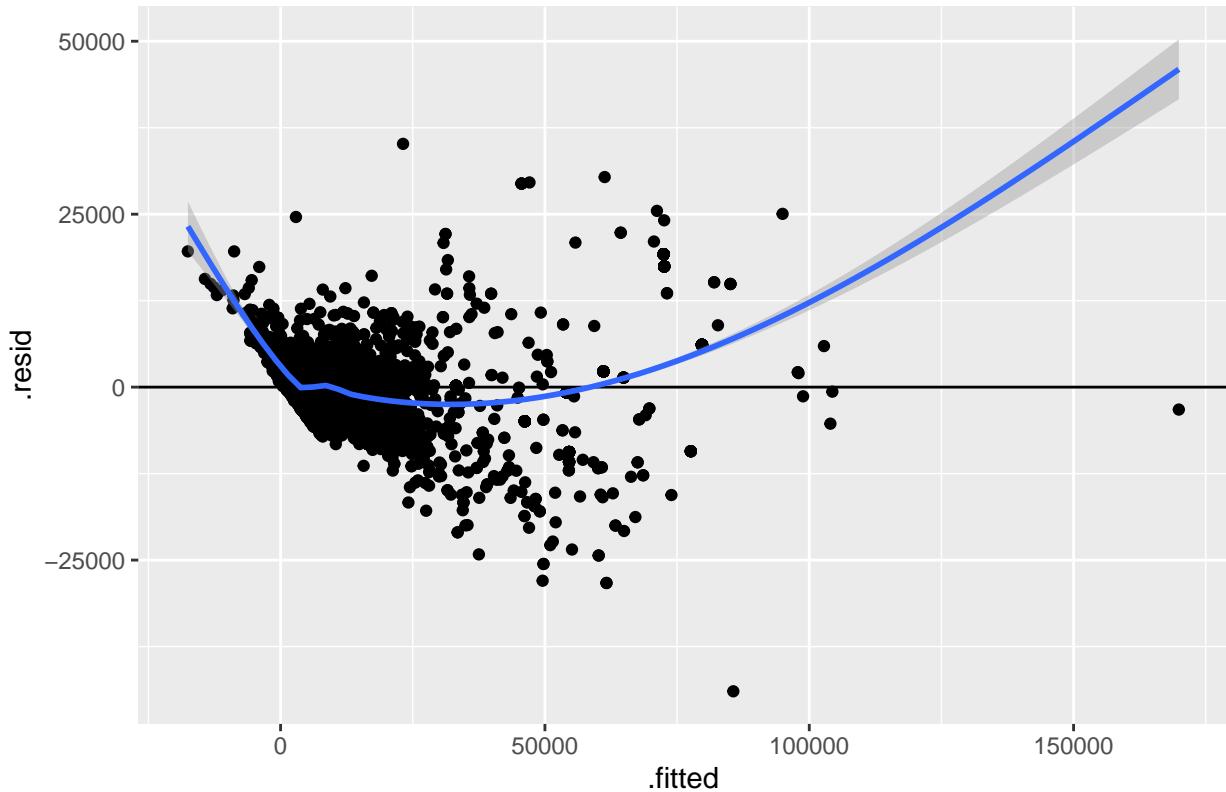
## 
## F Test for Heteroskedasticity
## -----
## Ho: Variance is homogenous
## Ha: Variance is not homogenous
##
## Variables: fitted values of selling_price
##
##      Test Summary
## -----
## Num DF     =     1
## Den DF     =    7817
## F          =   2747.151
## Prob > F  =    0.000

ggplot(new_Interac_firstordermodel, aes(x=.fitted, y=.resid)) +geom_point()+
  geom_hline(yintercept = 0) +geom_smooth()+
  ggtitle("Residual plot: Residual vs Fitted values")

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

```

Residual plot: Residual vs Fitted values



From the result of the Breusch-Pagan test, since the p-value:  $2.2e-16 < 0.05$ , we reject the null hypothesis and conclude that heteroscedasticity is present in the model.

### Checking the normality assumption

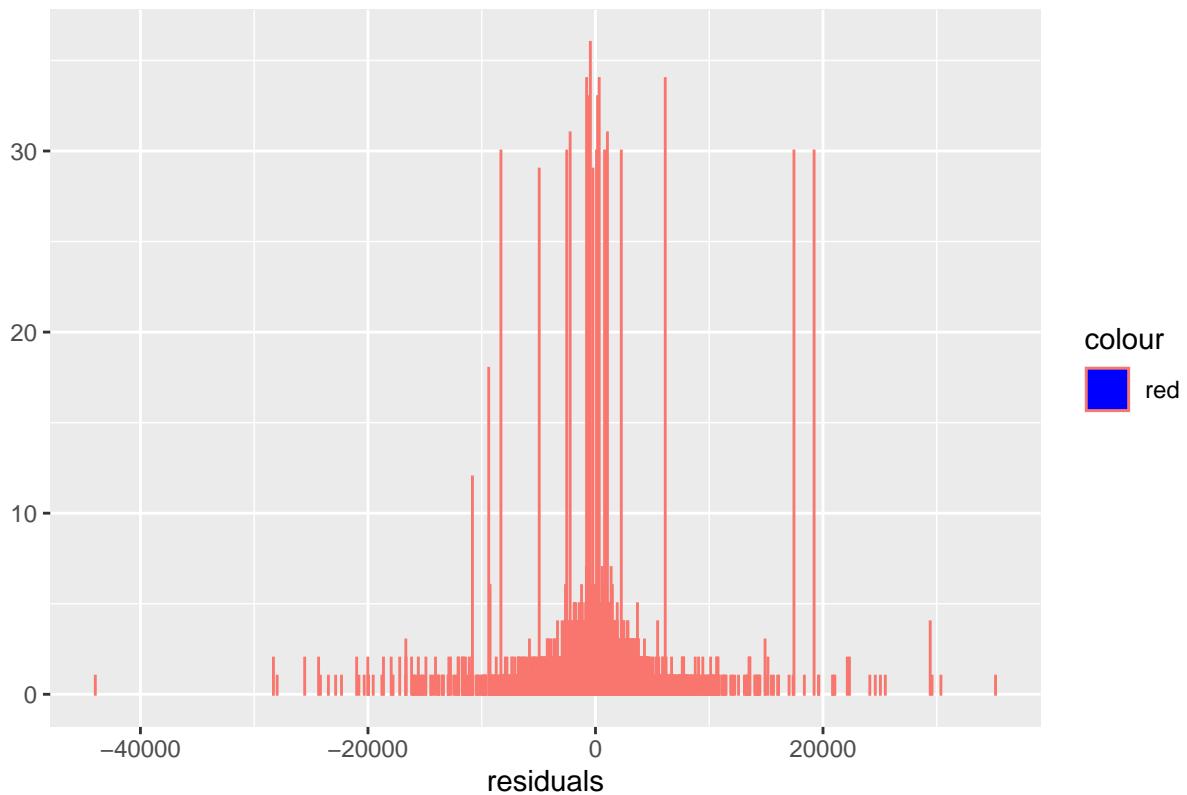
$H_0$ : the sample data are significantly normally distributed

$H_{alpha}$ : the sample data are not significantly normally distributed

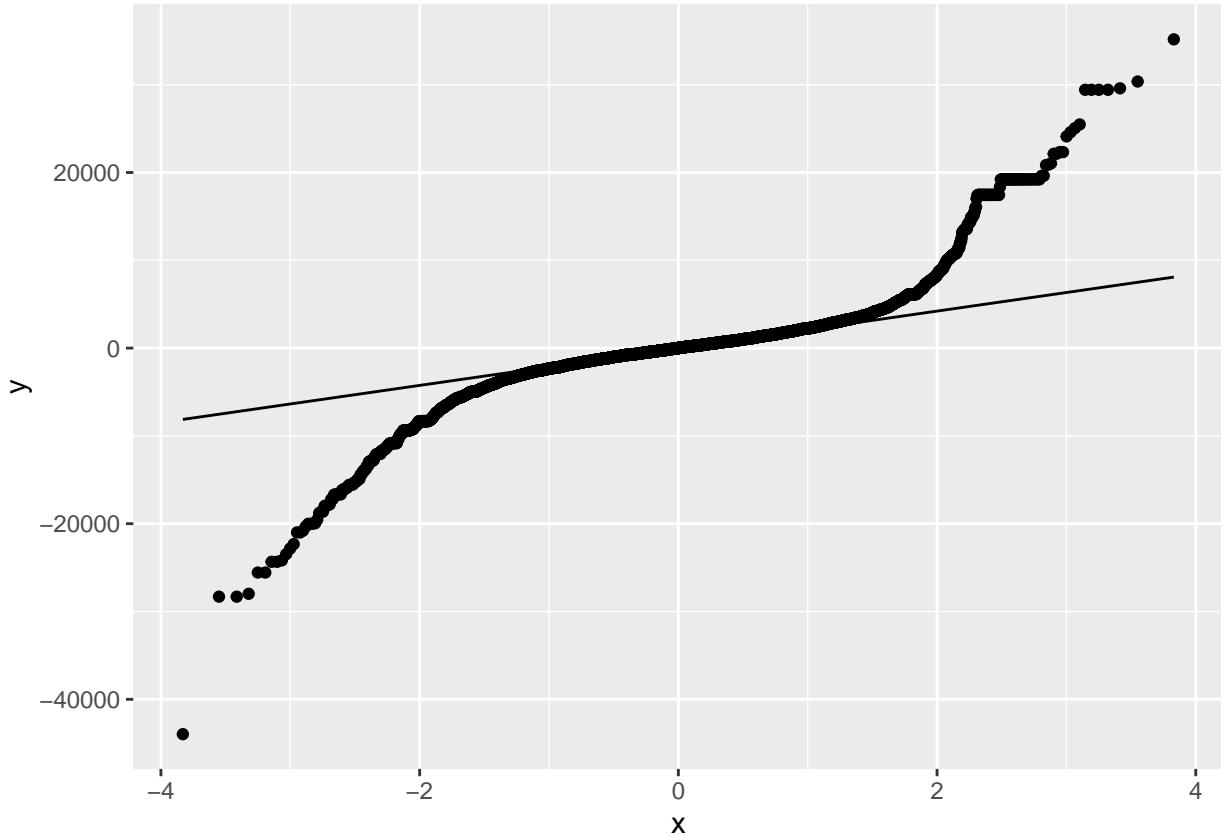
#### #Testing for Normality

```
qplot(residuals(new_Interac_firstordermodel) ,
      geom = "histogram",
      binwidth = 1,
      main = "Histogram of residual",
      xlab = "residuals",color = "red",
      fill = I("blue"))
```

Histogram of residual



```
#Q-Q plot
ggplot(Car1,aes(sample = new_Interac_firstordermodel$residuals))+  
  stat_qq()+ stat_qq_line()
```



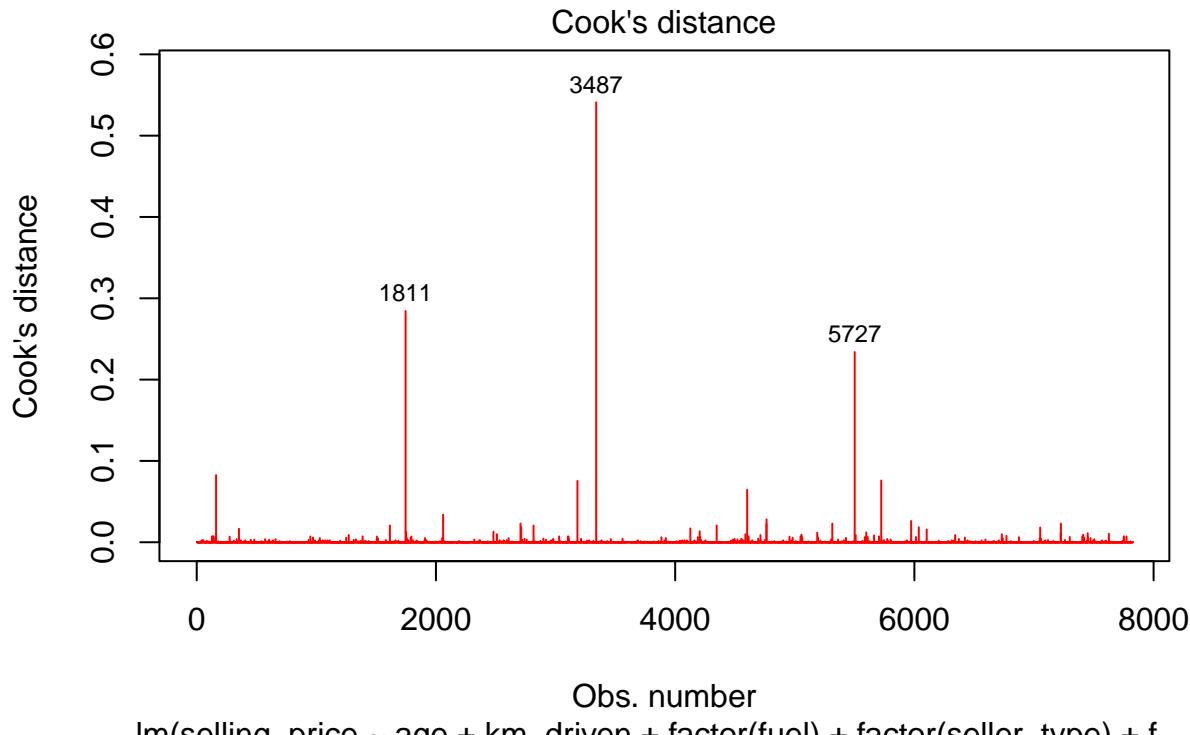
From the plot of the histogram of residuals and that of the normal q-q plot, there is a lot of deviation at the upper and lower tail of the plot. Hence, we conclude that the model is not normally distributed.

### 13. Check for outliers using the cook's distance.

```
Car1[cooks.distance(new_Interac_firstordermodel)>0.5,] #have Cook
  ↵ statistics larger than 0.1
```

```
##      age selling_price km_driven   fuel seller_type transmission         owner
## 3487    13        9166.667    2360457 Petrol Individual       Manual Second Owner
## NA      NA             NA          NA <NA>           <NA>           <NA>           <NA>
## NA.1    NA             NA          NA <NA>           <NA>           <NA>           <NA>
##      mileage engine max_power seats
## 3487     18.6    1197      81.83     5
## NA        NA      NA          NA      NA
## NA.1     NA      NA          NA      NA
```

```
plot(new_Interac_firstordermodel,pch=18,col="red",which=c(4))
```



The plot of the Cook's distance shows there are 3 outliers in the dataset. However at a threshold distance of 0.5, there is only one influential point in the dataset.

#### 14. Model transformation

After getting a good model that predicts the selling price of a used car, the model was tested for linearity, equal variance and normality assumptions and it failed all the test. Therefore we conducted a box cox transformation on the model in order to improve it.

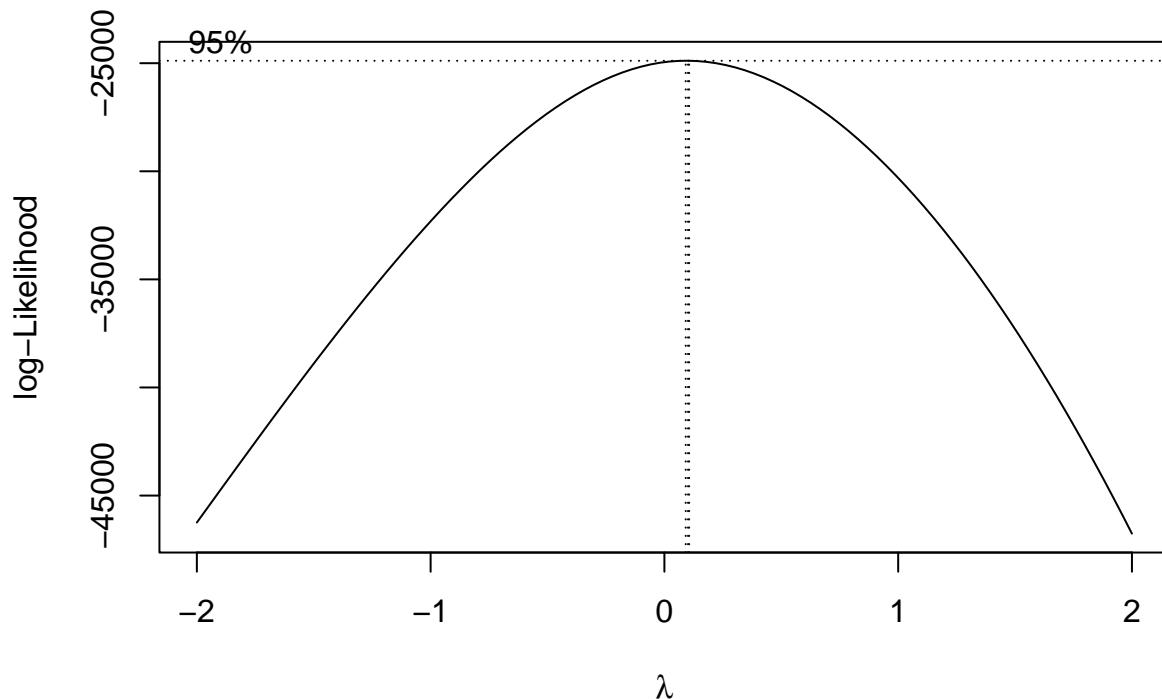
```
library(MASS)

## 
## Attaching package: 'MASS'

## The following object is masked from 'package:olsrr':
## 
##     cement

## The following object is masked from 'package:dplyr':
## 
##     select
```

```
bc = boxcox(new_Interac_firstordermodel, lambda = seq(-2,2))
```



```
bestlambda=bc$x[which(bc$y==max(bc$y))]
bestlambda
```

```
## [1] 0.1010101
```

```
bcmmodel=lm(((selling_price^0.1010101)-1)/0.1010101) ~age + km_driven +
  ↪ factor(fuel) + factor(seller_type)
  + factor(transmission)+ factor(owner)+ mileage +
  ↪ engine + max_power + age*km_driven +
  ↪ age*factor(fuel) + age*factor(seller_type)
  + age*factor(transmission) + age*factor(owner) +
  ↪ age*engine + age*max_power +
  ↪ km_driven*factor(seller_type)
  + km_driven*factor(transmission) +
  ↪ km_driven*factor(transmission) +
  ↪ km_driven*max_power +
  ↪ factor(fuel)*factor(transmission)
  + factor(fuel)*factor(owner) + factor(fuel)*mileage +
  ↪ factor(fuel)*engine + factor(fuel)*max_power +
  ↪ factor(seller_type)*mileage
```

```

+ factor(seller_type)*engine +
  ↵ factor(seller_type)*max_power +
  ↵ factor(transmission)*factor(owner) +
  ↵ factor(transmission)*engine
+ factor(transmission)*max_power +
  ↵ factor(owner)*mileage + factor(owner)*engine +
  ↵ mileage*engine + mileage*max_power
+ engine*max_power
,data = Car1)
summary(bcmodel)

## Call:
## lm(formula = (((selling_price^0.1010101) - 1)/0.1010101) ~ age +
##   km_driven + factor(fuel) + factor(seller_type) + factor(transmission) +
##   factor(owner) + mileage + engine + max_power + age * km_driven +
##   age * factor(fuel) + age * factor(seller_type) + age * factor(transmission) +
##   age * factor(owner) + age * engine + age * max_power + km_driven *
##   factor(seller_type) + km_driven * factor(transmission) +
##   km_driven * factor(transmission) + km_driven * max_power +
##   factor(fuel) * factor(transmission) + factor(fuel) * factor(owner) +
##   factor(fuel) * mileage + factor(fuel) * engine + factor(fuel) *
##   max_power + factor(seller_type) * mileage + factor(seller_type) *
##   engine + factor(seller_type) * max_power + factor(transmission) *
##   factor(owner) + factor(transmission) * engine + factor(transmission) *
##   max_power + factor(owner) * mileage + factor(owner) * engine +
##   mileage * engine + mileage * max_power + engine * max_power,
##   data = Car1)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -3.5848 -0.3732  0.0464  0.4050  3.4017
##
## Coefficients: (3 not defined because of singularities)
##              Estimate
## (Intercept) 9.262e+00
## age          -2.111e-01
## km_driven   -9.230e-06
## factor(fuel)Petrol 1.974e-01
## factor(seller_type)Individual 2.977e-01
## factor(seller_type)Trustmark Dealer -7.240e-01
## factor(transmission)Manual 6.202e-02
## factor(owner)Fourth & Above Owner 8.247e-01
## factor(owner)Second Owner -6.375e-01

```

## factor(owner)Test Drive Car	2.217e+01
## factor(owner)Third Owner	3.331e-01
## mileage	1.206e-01
## engine	2.174e-03
## max_power	5.538e-02
## age:km_driven	3.757e-07
## age:factor(fuel)Petrol	5.708e-02
## age:factor(seller_type)Individual	8.253e-03
## age:factor(seller_type)Trustmark Dealer	-1.342e-01
## age:factor(transmission)Manual	-1.390e-02
## age:factor(owner)Fourth & Above Owner	3.314e-02
## age:factor(owner)Second Owner	-2.528e-03
## age:factor(owner)Test Drive Car	NA
## age:factor(owner)Third Owner	-3.713e-04
## age:engine	2.260e-05
## age:max_power	-1.652e-03
## km_driven:factor(seller_type)Individual	2.842e-07
## km_driven:factor(seller_type)Trustmark Dealer	1.358e-05
## km_driven:factor(transmission)Manual	4.739e-06
## km_driven:max_power	6.811e-09
## factor(fuel)Petrol:factor(transmission)Manual	2.577e-01
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	-6.804e-01
## factor(fuel)Petrol:factor(owner)Second Owner	-3.029e-02
## factor(fuel)Petrol:factor(owner)Test Drive Car	-5.499e+00
## factor(fuel)Petrol:factor(owner)Third Owner	-2.341e-01
## factor(fuel)Petrol:mileage	-4.287e-02
## factor(fuel)Petrol:engine	-1.724e-04
## factor(fuel)Petrol:max_power	1.219e-04
## factor(seller_type)Individual:mileage	-1.085e-02
## factor(seller_type)Trustmark Dealer:mileage	-1.125e-02
## factor(seller_type)Individual:engine	1.848e-04
## factor(seller_type)Trustmark Dealer:engine	5.496e-04
## factor(seller_type)Individual:max_power	-5.570e-03
## factor(seller_type)Trustmark Dealer:max_power	2.891e-03
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	-6.039e-01
## factor(transmission)Manual:factor(owner)Second Owner	-1.473e-01
## factor(transmission)Manual:factor(owner)Test Drive Car	NA
## factor(transmission)Manual:factor(owner)Third Owner	-1.939e-01
## factor(transmission)Manual:engine	1.408e-04
## factor(transmission)Manual:max_power	-8.596e-03
## factor(owner)Fourth & Above Owner:mileage	-5.163e-03
## factor(owner)Second Owner:mileage	2.397e-02
## factor(owner)Test Drive Car:mileage	-9.889e-01
## factor(owner)Third Owner:mileage	-1.073e-02
## factor(owner)Fourth & Above Owner:engine	-3.755e-04

## factor(owner)Second Owner:engine		1.445e-04
## factor(owner)Test Drive Car:engine		NA
## factor(owner)Third Owner:engine		-6.380e-05
## mileage:engine		-7.069e-05
## mileage:max_power		2.808e-04
## engine:max_power		-7.978e-06
##	Std. Error	t value
## (Intercept)	4.681e-01	19.787
## age	2.159e-02	-9.776
## km_driven	1.322e-06	-6.980
## factor(fuel)Petrol	2.505e-01	0.788
## factor(seller_type)Individual	2.246e-01	1.325
## factor(seller_type)Trustmark Dealer	5.631e-01	-1.286
## factor(transmission)Manual	1.619e-01	0.383
## factor(owner)Fourth & Above Owner	8.187e-01	1.007
## factor(owner)Second Owner	2.922e-01	-2.182
## factor(owner)Test Drive Car	1.384e+01	1.602
## factor(owner)Third Owner	5.353e-01	0.622
## mileage	1.420e-02	8.495
## engine	2.271e-04	9.572
## max_power	2.540e-03	21.798
## age:km_driven	4.233e-08	8.876
## age:factor(fuel)Petrol	7.334e-03	7.783
## age:factor(seller_type)Individual	1.033e-02	0.799
## age:factor(seller_type)Trustmark Dealer	5.717e-02	-2.348
## age:factor(transmission)Manual	1.258e-02	-1.104
## age:factor(owner)Fourth & Above Owner	1.636e-02	2.026
## age:factor(owner)Second Owner	6.168e-03	-0.410
## age:factor(owner)Test Drive Car		NA NA
## age:factor(owner)Third Owner	1.071e-02	-0.035
## age:engine	8.597e-06	2.629
## age:max_power	1.109e-04	-14.896
## km_driven:factor(seller_type)Individual	8.926e-07	0.318
## km_driven:factor(seller_type)Trustmark Dealer	4.877e-06	2.785
## km_driven:factor(transmission)Manual	9.774e-07	4.849
## km_driven:max_power	5.644e-09	1.207
## factor(fuel)Petrol:factor(transmission)Manual	7.926e-02	3.251
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner	1.727e-01	-3.941
## factor(fuel)Petrol:factor(owner)Second Owner	5.829e-02	-0.520
## factor(fuel)Petrol:factor(owner)Test Drive Car	4.220e+00	-1.303
## factor(fuel)Petrol:factor(owner)Third Owner	9.625e-02	-2.433
## factor(fuel)Petrol:mileage	7.150e-03	-5.996
## factor(fuel)Petrol:engine	1.321e-04	-1.305
## factor(fuel)Petrol:max_power	1.438e-03	0.085
## factor(seller_type)Individual:mileage	7.132e-03	-1.521

## factor(seller_type)Trustmark Dealer:mileage	2.320e-02	-0.485
## factor(seller_type)Individual:engine	9.879e-05	1.871
## factor(seller_type)Trustmark Dealer:engine	2.228e-04	2.466
## factor(seller_type)Individual:max_power	1.081e-03	-5.151
## factor(seller_type)Trustmark Dealer:max_power	4.292e-03	0.674
## factor(transmission)Manual:factor(owner)Fourth & Above Owner	2.890e-01	-2.089
## factor(transmission)Manual:factor(owner)Second Owner	7.998e-02	-1.842
## factor(transmission)Manual:factor(owner)Test Drive Car	NA	NA
## factor(transmission)Manual:factor(owner)Third Owner	1.767e-01	-1.097
## factor(transmission)Manual:engine	1.090e-04	1.291
## factor(transmission)Manual:max_power	1.193e-03	-7.204
## factor(owner)Fourth & Above Owner:mileage	2.304e-02	-0.224
## factor(owner)Second Owner:mileage	7.942e-03	3.018
## factor(owner)Test Drive Car:mileage	6.208e-01	-1.593
## factor(owner)Third Owner:mileage	1.529e-02	-0.702
## factor(owner)Fourth & Above Owner:engine	1.882e-04	-1.996
## factor(owner)Second Owner:engine	7.174e-05	2.014
## factor(owner)Test Drive Car:engine	NA	NA
## factor(owner)Third Owner:engine	1.245e-04	-0.513
## mileage:engine	8.894e-06	-7.948
## mileage:max_power	8.372e-05	3.354
## engine:max_power	6.439e-07	-12.390
##	Pr(> t )	
## (Intercept)	< 2e-16 ***	
## age	< 2e-16 ***	
## km_driven	3.20e-12 ***	
## factor(fuel)Petrol	0.430618	
## factor(seller_type)Individual	0.185146	
## factor(seller_type)Trustmark Dealer	0.198600	
## factor(transmission)Manual	0.701587	
## factor(owner)Fourth & Above Owner	0.313804	
## factor(owner)Second Owner	0.029141 *	
## factor(owner)Test Drive Car	0.109177	
## factor(owner)Third Owner	0.533799	
## mileage	< 2e-16 ***	
## engine	< 2e-16 ***	
## max_power	< 2e-16 ***	
## age:km_driven	< 2e-16 ***	
## age:factor(fuel)Petrol	8.00e-15 ***	
## age:factor(seller_type)Individual	0.424230	
## age:factor(seller_type)Trustmark Dealer	0.018911 *	
## age:factor(transmission)Manual	0.269445	
## age:factor(owner)Fourth & Above Owner	0.042822 *	
## age:factor(owner)Second Owner	0.681905	
## age:factor(owner)Test Drive Car	NA	

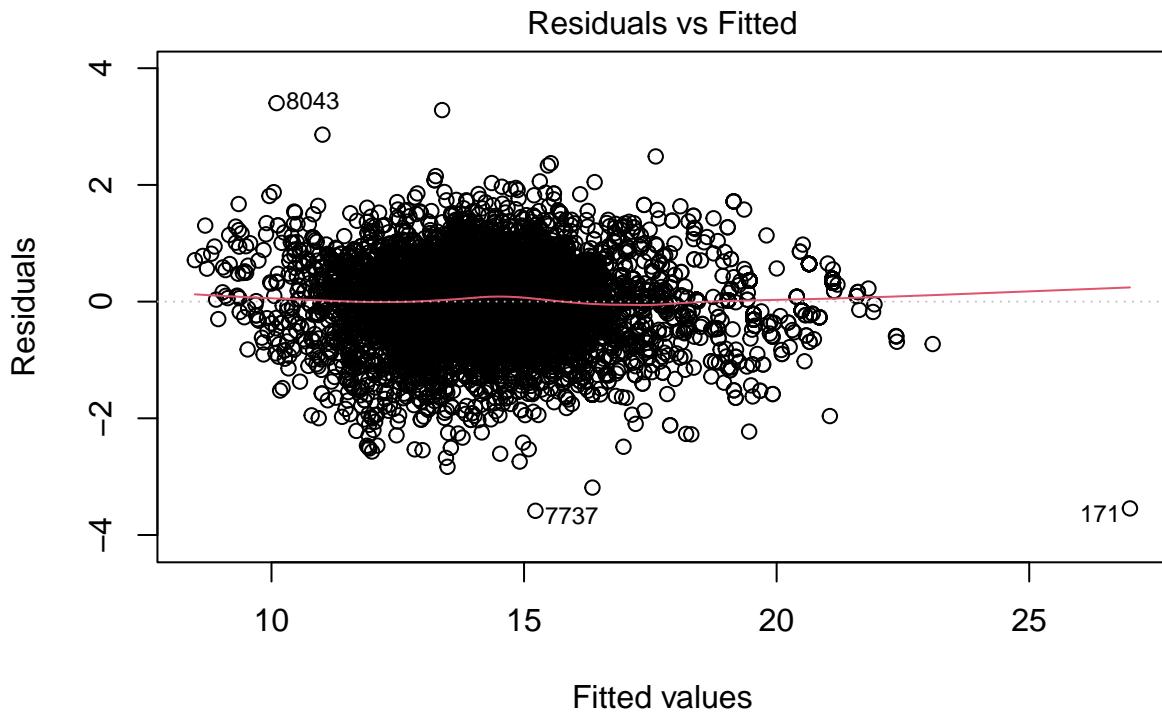
```

## age:factor(owner)Third Owner          0.972349
## age:engine                          0.008586 **
## age:max_power                      < 2e-16 ***
## km_driven:factor(seller_type)Individual 0.750187
## km_driven:factor(seller_type)Trustmark Dealer 0.005359 **
## km_driven:factor(transmission)Manual    1.27e-06 ***
## km_driven:max_power                 0.227567
## factor(fuel)Petrol:factor(transmission)Manual 0.001154 **
## factor(fuel)Petrol:factor(owner)Fourth & Above Owner 8.20e-05 ***
## factor(fuel)Petrol:factor(owner)Second Owner 0.603282
## factor(fuel)Petrol:factor(owner)Test Drive Car 0.192525
## factor(fuel)Petrol:factor(owner)Third Owner 0.015008 *
## factor(fuel)Petrol:mileage            2.11e-09 ***
## factor(fuel)Petrol:engine            0.191769
## factor(fuel)Petrol:max_power        0.932461
## factor(seller_type)Individual:mileage 0.128250
## factor(seller_type)Trustmark Dealer:mileage 0.627936
## factor(seller_type)Individual:engine 0.061404 .
## factor(seller_type)Trustmark Dealer:engine 0.013669 *
## factor(seller_type)Individual:max_power 2.65e-07 ***
## factor(seller_type)Trustmark Dealer:max_power 0.500554
## factor(transmission)Manual:factor(owner)Fourth & Above Owner 0.036715 *
## factor(transmission)Manual:factor(owner)Second Owner 0.065556 .
## factor(transmission)Manual:factor(owner)Test Drive Car NA
## factor(transmission)Manual:factor(owner)Third Owner 0.272516
## factor(transmission)Manual:engine     0.196764
## factor(transmission)Manual:max_power 6.37e-13 ***
## factor(owner)Fourth & Above Owner:mileage 0.822693
## factor(owner)Second Owner:mileage    0.002556 **
## factor(owner)Test Drive Car:mileage 0.111194
## factor(owner)Third Owner:mileage    0.482761
## factor(owner)Fourth & Above Owner:engine 0.045997 *
## factor(owner)Second Owner:engine    0.044057 *
## factor(owner)Test Drive Car:engine NA
## factor(owner)Third Owner:engine    0.608248
## mileage:engine                     2.16e-15 ***
## mileage:max_power                  0.000799 ***
## engine:max_power                  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6632 on 7762 degrees of freedom
## Multiple R-squared:  0.8985, Adjusted R-squared:  0.8978
## F-statistic: 1227 on 56 and 7762 DF,  p-value: < 2.2e-16

```

Checking the linearity assumption after conducting a boxcox transformation

```
par(mfrow=c(1,1))
plot(bcmodel, which = 1)
```



```
lm(((selling_price^0.1010101) - 1)/0.1010101) ~ age + km_driven + factor(f ...
```

After conducting a box cox transformation on the model, no discernible pattern appears in the residual vs fitted plot. Indicating that the linearity assumption of the model is valid.

Checking the equal variance assumption

$H_0$ : heteroscedasticity is not present(homoscedasticity)

$H_{alpha}$ : heteroscedasticity is present

```
library(lmtest)
bpptest(bcmodel)
```

```
##
## studentized Breusch-Pagan test
##
## data: bcmodel
## BP = 903.1, df = 56, p-value < 2.2e-16
```

```

library(olsrr)
ols_test_f(bcmodel)

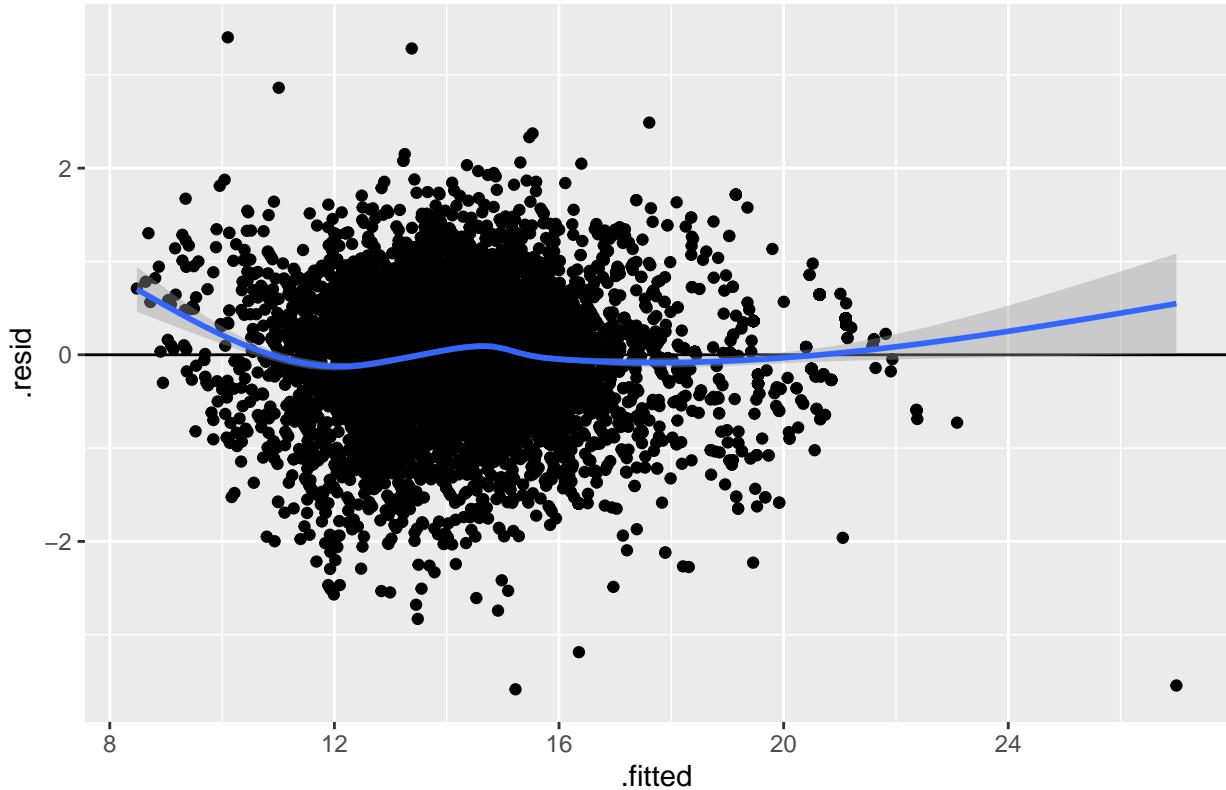
## F Test for Heteroskedasticity
## -----
## Ho: Variance is homogenous
## Ha: Variance is not homogenous
##
## Variables: fitted values of (((selling_price^0.1010101) - 1)/0.1010101)
##
##          Test Summary
## -----
## Num DF      =     1
## Den DF      =    7817
## F            =   17.20734
## Prob > F    =  3.386882e-05

ggplot(bcmodel, aes(x=.fitted, y=.resid)) +geom_point()+
  geom_hline(yintercept = 0) +geom_smooth()+
  ggtitle("Residual plot: Residual vs Fitted values")

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

```

## Residual plot: Residual vs Fitted values



Although a box cox transformation was conducted on the model, the homoscedasticity assumption was failed to be met, since the p-value of the Breusch-Pagan test:  $2.2e-16 < 0.05$ . Indicating we reject the null hypothesis.

### Checking the normality assumption

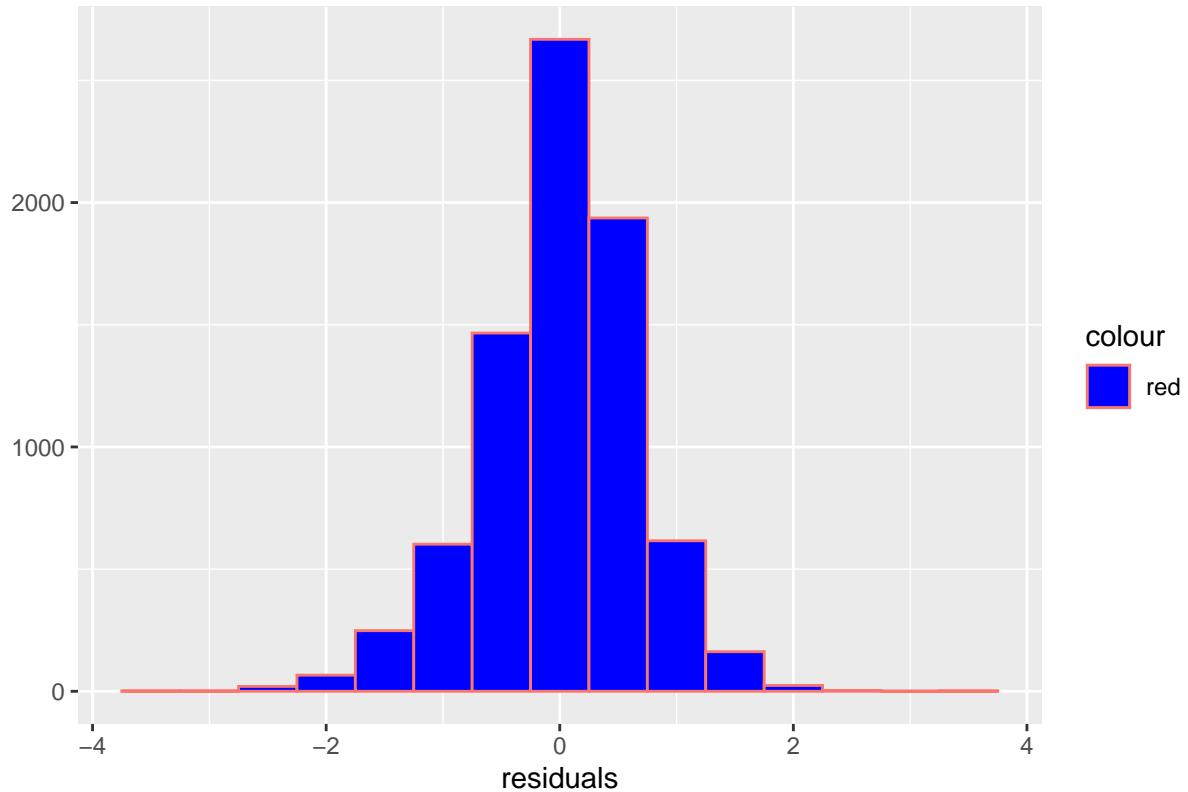
$H_0$ : the sample data are significantly normally distributed

$H_{alpha}$ : the sample data are not significantly normally distributed

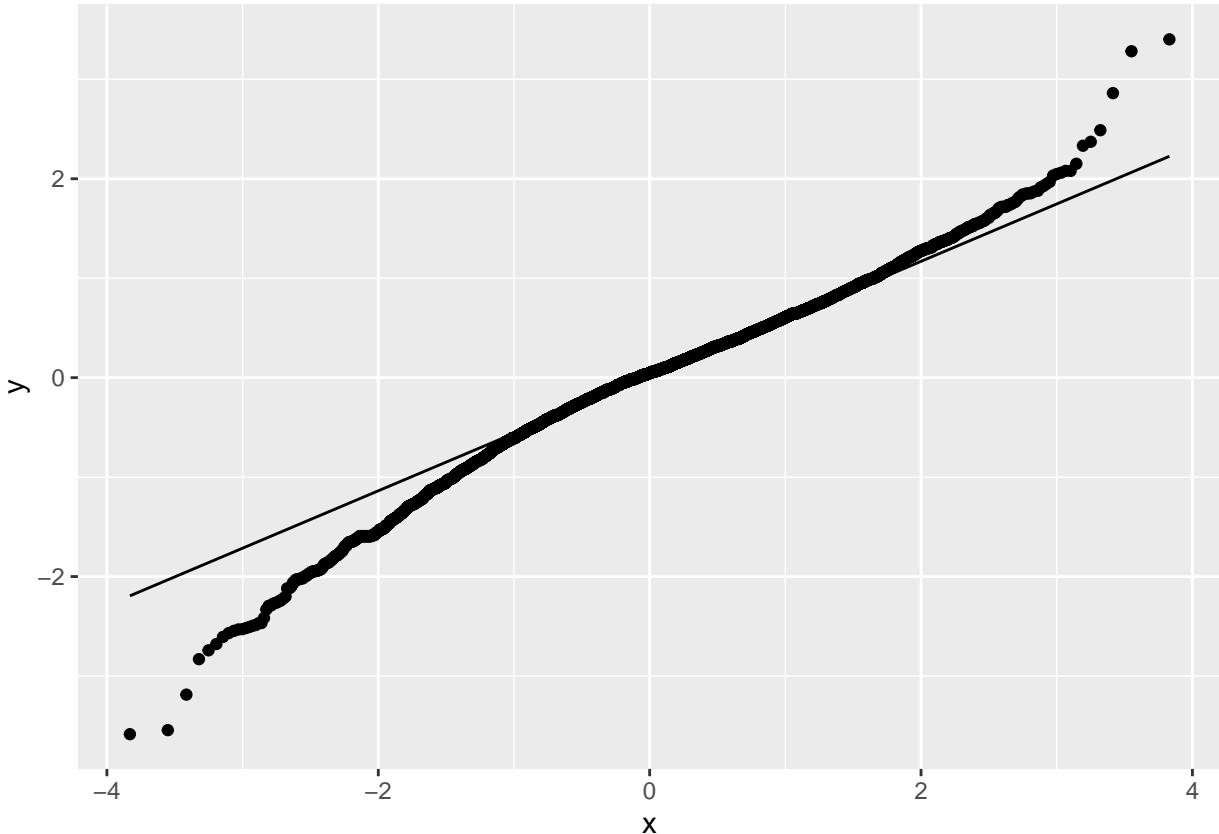
#### #Testing for Normality

```
qplot(residuals(bcmodel),
      geom = "histogram",
      binwidth = 0.5,
      main = "Histogram of residual",
      xlab = "residuals", color = "red",
      fill = I("blue"))
```

Histogram of residual



```
#Q-Q plot
ggplot(Car1,aes(sample = bcmodel$residuals))+  
  stat_qq()+ stat_qq_line()
```



From the plot of the histogram and residuals, the bars form a bell shape indicating the model is approximately normally distributed. This is also supported by the normal Q-Q plot.

After checking the linearity, homoscedasticity, normality assumption of the best interaction model, we discovered the model failed to satisfy all the model assumptions. Therefore, we conducted a box cox transformation on the model. The resulting model has an  $R^2_{adj}$  value of 0.8978. Indicating 89.78% of the variation of the selling price of a used car can be explained by the model.

Finally, although the  $R^2_{adj}$  value of the best interaction model: 0.9114 is slightly higher than that of the boxcox transformation model: 0.8978, we chose the boxcox transformation model as the best model for predicting the selling price of a used car, since it satisfies the model assumptions.

## Testing the Model

### 15. Using the Final Model to Predict the Selling Price of a Used Car

```
newdata = data.frame(age=0, km_driven=10, fuel="Petrol", seller_type
                     ="Individual", transmission="Automatic", owner="First Owner",
                     mileage=15.4, engine=1400, max_power=100)
predict(bcmodel,newdata,interval="predict")
```

```

##      fit      lwr      upr
## 1 16.63265 15.32724 17.93806

```

Finally, after designing and testing the model, we used put it to test, to predict the selling price of a used car with:

## Conclusion

1. The best fit model equation is stated below:

$$\begin{aligned}
\hat{\text{sellingprice}} = & 9.262 - 0.2111 \text{ age} - 9.23e-06 \text{ km\_driven} + 1.974e-01 \text{ fuel} + 2.977e-01 \\
& \text{seller\_type(individual)} - 7.240e-01 \text{ seller\_type(Trustmark Dealer)} + 6.202e-02 \text{ Transmission(manual)} \\
& + 8.247e-01 \text{ owner(fourth \& above)} - 6.375e-01 \text{ owner(second)} + 2.217e+01 \\
& \text{owner(test drive)} + 3.331e-01 \text{ owner(third)} + 1.206e-01 \text{ mileage} + 2.174e-03 \text{ engine} \\
& + 5.538e-02 \text{ max\_power} + 3.757e-07 \text{ age} * \text{km\_driven} + 5.708e-02 \text{ age} * \text{fuel(petrol)} \\
& + 8.253e-03 \text{ age} * \text{seller\_type(Individual)} - 1.342e-01 \text{ age} * \text{seller\_type(Trustmark Dealer)} \\
& - 1.390e-02 \text{ age} * \text{transmission(Manual)} + 3.314e-02 \text{ age} * \text{owner(Fourth \& Above Owner)} \\
& - 2.528e-03 \text{ age} * \text{owner(Second Owner)} - 3.713e-04 \text{ age} * \text{owner(Third owner)} + 2.260e-05 \\
& \text{age} * \text{engine} - 1.652e-03 \text{ age} * \text{max\_power} + 2.842e-07 \text{ km\_driven} * \text{seller\_type(Individual)} \\
& + 1.358e-05 \text{ km\_driven} * \text{seller\_type(Trustmark Dealer)} + 4.739e-06 \text{ km\_driven} * \\
& \text{transmission(Manual)} + 6.811e-09 \text{ km\_driven} * \text{max\_power} + 2.577e-01 \text{ fuel(Petrol)} * \\
& \text{transmission (Manual)} - 6.804e-01 \text{ fuel(Petrol)} * \text{owner(Fourth \& Above Owner)} - 3.029e-02 \\
& \text{fuel Petrol} * \text{owner(second)} - 5.499 \text{ fuel(Petrol)} * \text{owner(test drive)} - 2.341e-01 \text{ fuel(Petrol)} \\
& * \text{owner(third)} - 4.287e-02 \text{ fuel(petrol)} * \text{mileage} - 1.724e-04 \text{ fuel(petrol)} * \text{engine} + 1.219e-04 \\
& \text{fuel(petrol)} * \text{max\_power} - 1.085e-02 \text{ seller\_type(Individual)} * \text{mileage} - 1.125e-02 \\
& \text{seller\_type(Trustmark Dealer)} * \text{mileage} + 1.848e-04 \text{ seller\_type(Individual)} * \text{engine} + \\
& 5.496e-04 \text{ seller\_type(Trustmark Dealer)} * \text{engine} - 5.570e-03 \text{ seller\_type(Individual)} * \\
& \text{max\_power} + 2.891e-03 \text{ seller\_type (Trustmark Dealer)} * \text{max\_power} - 6.039e-01 \\
& \text{transmission(Manual)} * \text{owner(Fourth \& Above Owner)} - 1.473e-01 \text{ transmission(Manual)} * \\
& \text{owner(Second Owner)} \\
& - 1.939e-01 \text{ transmission(Manual)} * \text{owner(Third Owner)} \\
& + 1.408e-04 \text{ transmission(Manual)} * \text{engine} - 8.596e-03 \text{ transmission(Manual)} * \text{max\_power} \\
& - 5.163e-03 \text{ owner(Fourth \& Above Owner)} * \text{mileage} \\
& + 2.397e-02 \text{ owner(Second Owner)} * \text{mileage} \\
& - 9.889e-01 \text{ owner(Test Drive Car)} * \text{mileage} \\
& - 1.073e-02 \text{ owner(Third Owner)} * \text{mileage} \\
& - 3.755e-04 \text{ owner(Fourth \& Above Owner)} * \text{engine} \\
& 1.445e-04 \text{ owner(Second Owner)} * \text{engine} \\
& - 6.380e-05 \text{ owner(Third Owner)} * \text{engine} \\
& - 7.069e-05 \text{ mileage} * \text{engine} \\
& + 2.808e-04 \text{ mileage} * \text{max\_power} \\
& - 7.978e-06 \text{ engine} * \text{max\_power}
\end{aligned}$$

- After checking the linearity, homoscedasticity, normality assumption of the best interaction model, we discovered the model failed to satisfy all the model assumptions. Therefore, we conducted a box cox transformation on the model. The resulting model has an  $R^2_{adj}$  value of 0.8978. Indicating 89.78% of the variation of the selling price of a used car can be explained by the model.

Finally, although the  $R^2_{adj}$  value of the best interaction model: 0.9114 is slightly higher than that of the boxcox transformation model: 0.8978, we chose the boxcox transformation model as the best model for predicting the selling price of a used car, since it satisfies the model assumptions.

- The significant independent variables that are used to predict the selling\_price are as follows: age km\_driven fuel\_seller\_type Transmission owner mileage engine max\_power

## Discussion

- Although the result of the partial F-test conducted above rejects the null hypothesis, indicating the number of seats variable be kept in the model, but by conducting the backward elimination, the best subset and a partial t-test, we concluded on using the model without the seat predictor variable.
- From our analysis, we thought that the best fit model would be the interaction model. But upon checking the linearity, homoscedasticity, normality assumption we found that in-spite of having the  $R^2_{adj}$  a little less than the interaction model, the boxcox transformation model was a better fit. But even in this model, we find that heteroscedasticity still exists.
- To further better the existing model, we can try to remove the heteroscedasticity that is present. We can aslo consider other independent variables to predict the model. Other machine learning models can be implemented to get better results.

## References

- Vehicle dataset | By Anon | Container: kaggle.com URL: <https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho?select=Car+details+v3.csv> | Accessed on: 2021-11-20