Statistics Project

Amirabbas Afzali - 400100662

At first , we load data :

```
my_dataset = read.csv('D:/Term4/Statistics/R/Project/CarPrice_Assignment.csv')
```

```
head(my_dataset)
```

```
##   X car_ID symboling                  CarName fueltype aspiration doornumber
## 1 0      1         3       alfa-romero giulia      gas        std        two
## 2 1      2         3      alfa-romero stelvio      gas        std        two
## 3 2      3         1 alfa-romero Quadrifoglio      gas        std        two
## 4 3      4         2              audi 100 ls      gas        std       four
## 5 4      5         2               audi 100ls      gas        std       four
## 6 5      6         2                 audi fox      gas        std        two
##       carbody drivewheel enginelocation wheelbase carlength carwidth carheight
## 1 convertible        rwd          front      88.6     168.8     64.1      48.8
## 2 convertible        rwd          front      88.6     168.8     64.1      48.8
## 3   hatchback        rwd          front      94.5     171.2     65.5      52.4
## 4       sedan        fwd          front      99.8     176.6     66.2      54.3
## 5       sedan        4wd          front      99.4     176.6     66.4      54.3
## 6       sedan        fwd          front      99.8     177.3     66.3      53.1
##   curbweight enginetype cylindernumber enginesize fuelsystem boreratio stroke
## 1       2548       dohc           four        130       mpfi      3.47   2.68
## 2       2548       dohc           four        130       mpfi      3.47   2.68
## 3       2823       ohcv                       152       mpfi      2.68   3.47
## 4       2337        ohc           four        109       mpfi      3.19   3.40
## 5         NA        ohc                       136       mpfi      3.19   3.40
## 6       2507        ohc           five        136       mpfi        NA   3.40
##   compressionratio horsepower peakrpm citympg highwaympg price
## 1              9.0        111    5000      21         27 13495
## 2              9.0        111    5000      21         27 16500
## 3              9.0        154    5000      19         26 16500
## 4             10.0        102    5500      24         30 13950
## 5              8.0        115    5500      18         22 17450
## 6              8.5        110    5500      19         25 15250
```

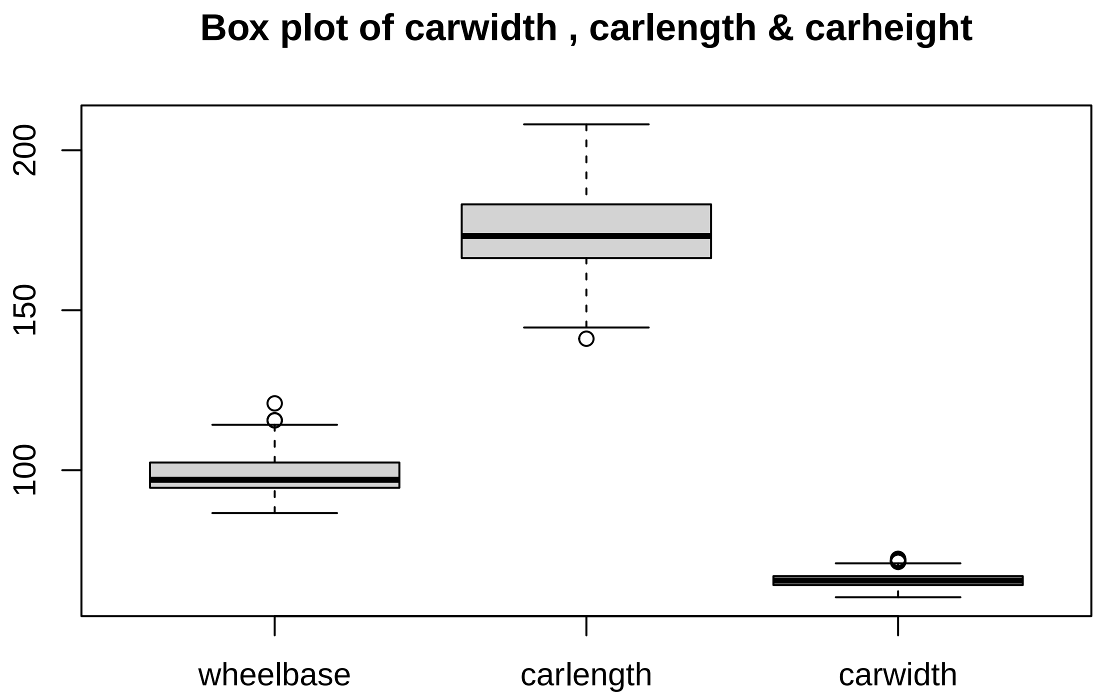
```
summary(my_dataset)
```

```
##        X            car_ID      symboling         CarName
##  Min.   :  0   Min.   :  1   Min.   :-2.0000   Length:205
##  1st Qu.: 51   1st Qu.: 52   1st Qu.: 0.0000   Class :character
##  Median :102   Median :103   Median : 1.0000   Mode  :character
##  Mean   :102   Mean   :103   Mean   : 0.8341
##  3rd Qu.:153   3rd Qu.:154   3rd Qu.: 2.0000
##  Max.   :204   Max.   :205   Max.   : 3.0000
##
##    fueltype          aspiration         doornumber          carbody
##  Length:205         Length:205         Length:205         Length:205
```

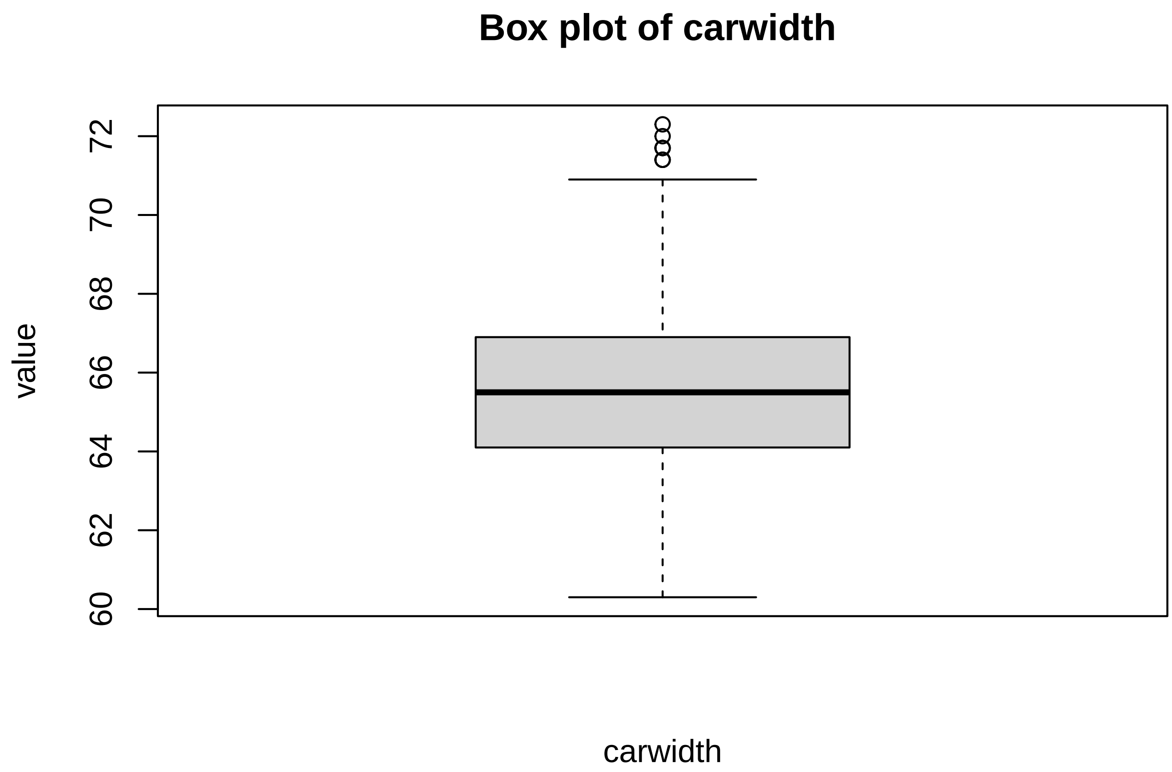
```
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   drivewheel        enginelocation      wheelbase       carlength
##  Length:205         Length:205         Min.   : 86.60   Min.   :141.1
##  Class :character   Class :character   1st Qu.: 94.50   1st Qu.:166.3
##  Mode  :character   Mode  :character   Median : 97.00   Median :173.2
##                                        Mean   : 98.76   Mean   :174.0
##                                        3rd Qu.:102.40   3rd Qu.:183.1
##                                        Max.   :120.90   Max.   :208.1
##
##     carwidth       carheight       curbweight    enginetype
##  Min.   :60.30   Min.   :47.80   Min.   :1488   Length:205
##  1st Qu.:64.10   1st Qu.:52.00   1st Qu.:2145   Class :character
##  Median :65.50   Median :54.10   Median :2417   Mode  :character
##  Mean   :65.91   Mean   :53.72   Mean   :2541
##  3rd Qu.:66.90   3rd Qu.:55.50   3rd Qu.:2919
##  Max.   :72.30   Max.   :59.80   Max.   :4066
##                                  NA's   :23
##  cylindernumber       enginesize     fuelsystem          boreratio
##  Length:205         Min.   : 61.0   Length:205         Min.   :2.540
##  Class :character   1st Qu.: 97.0   Class :character   1st Qu.:3.150
##  Mode  :character   Median :120.0   Mode  :character   Median :3.310
##                     Mean   :126.9                      Mean   :3.334
##                     3rd Qu.:141.0                      3rd Qu.:3.590
##                     Max.   :326.0                      Max.   :3.940
##                                                        NA's   :18
##      stroke      compressionratio   horsepower       peakrpm
##  Min.   :2.070   Min.   : 7.00    Min.   : 48.0   Min.   :4150
##  1st Qu.:3.110   1st Qu.: 8.60    1st Qu.: 70.0   1st Qu.:4800
##  Median :3.290   Median : 9.00    Median : 95.0   Median :5200
##  Mean   :3.255   Mean   :10.14    Mean   :104.1   Mean   :5125
##  3rd Qu.:3.410   3rd Qu.: 9.40    3rd Qu.:116.0   3rd Qu.:5500
##  Max.   :4.170   Max.   :23.00    Max.   :288.0   Max.   :6600
##
##     citympg        highwaympg        price
##  Min.   :13.00   Min.   :16.00   Min.   : 5118
##  1st Qu.:19.00   1st Qu.:25.00   1st Qu.: 7788
##  Median :24.00   Median :30.00   Median :10295
##  Mean   :25.22   Mean   :30.75   Mean   :13277
##  3rd Qu.:30.00   3rd Qu.:34.00   3rd Qu.:16503
##  Max.   :49.00   Max.   :54.00   Max.   :45400
##
```

Now we display the BOX plot related to *carlength*, *carwidth*, *carheight* and analyze them.

```
boxplot(my_dataset[,c("wheelbase","carlength","carwidth")],
        main = 'Box plot of carwidth , carlength & carheight')
```




```
boxplot(my_dataset[,"carwidth"],
        main = 'Box plot of carwidth '
        , xlab = 'carwidth'
        , ylab = 'value')
```



Data wrangling

```
cat('missing value of curbweight :', length(my_dataset$curbweight[is.na(my_dataset$curbweight)]))
```

```
## missing value of curbweight : 23
```

```
cat('\nmissing value of boreratio : ',length(my_dataset$boreratio[is.na(my_dataset$boreratio)]))
```

```
##
## missing value of boreratio :  18
```

```
cat('\nmissing value of carbody : ',length(my_dataset$carbody[nchar(my_dataset$carbody)==0]))
```

```
##
## missing value of carbody :  17
```

```
cat('\nmissing value of cylindernumber : ',length(my_dataset$cylindernumber[nchar(my_dataset$cylindernu
```

```
##
## missing value of cylindernumber :  20
```

Now we replace them :

```
my_dataset$boreratio[is.na(my_dataset$boreratio)] <- median(my_dataset$boreratio,na.rm = TRUE)
my_dataset$curbweight[is.na(my_dataset$curbweight)] <- median(my_dataset$curbweight,na.rm = TRUE)
my_dataset$cylindernumber[nchar(my_dataset$cylindernumber)==0] <- mode(my_dataset$cylindernumber)
my_dataset$carbody[nchar(my_dataset$carbody)==0] <- mode(my_dataset$carbody)
```

When replacing missing data in the **CarPrice_Assignment** dataset or any other dataset, the choice between using the median or the mean depends on the nature of the variable and the characteristics of the data. Here are some considerations to help you decide:

1. Outliers: If the variable in question is sensitive to outliers, it is generally more appropriate to use the median. The median is robust to outliers because it is not affected by extreme values. On the other hand, the mean can be influenced by outliers, especially if they are far from the central values of the distribution.

2. Skewness: Assess the skewness of the variable. If the data is skewed or not normally distributed, the median may be a better choice. The median is not affected by the skewness of the distribution, whereas the mean can be pulled towards the long tail of the distribution if it is skewed.

3. Data distributions: Consider the shape and characteristics of the data distribution. If the variable follows a symmetrical distribution with no significant outliers, the mean may provide a suitable estimate. However, if the data has a skewed or heavily tailed distribution, the median could provide a more representative estimate.

4. Impact on analysis: Think about the implications of using the median or the mean on downstream analysis. Depending on the specific analysis or modeling technique you plan to use, the mean or median might be more appropriate. For instance, linear regression models typically assume that the predictors have a linear relationship with the response. In these cases, imputing missing values with the mean might be more suitable.

The median is often used to replace missing data in data wrangling because it is a measure of central tendency that is less sensitive to extreme values or outliers compared to the mean. When dealing with missing data, it's important to handle it in a way that minimizes the impact on the overall analysis and maintains the integrity of the data.

Here are a few reasons why the median is commonly used for replacing missing data:

1. Robustness to outliers: The median is resistant to extreme values or outliers in the dataset. This makes it a suitable choice when the presence of outliers could distort the mean and affect imputation accuracy.

2. Preserving data distribution: The median reflects the value that separates the upper and lower halves of a dataset. By using the median to replace missing values, the general distribution and order of the data are preserved to a certain extent.

3. Non-parametric estimation: Unlike the mean, the median does not make any assumptions about the underlying data distribution. This makes it a more robust measure when the data does not follow a normal distribution or when we have limited information about the data characteristics.

Regarding the difference between mean and median, it relates to their calculation and sensitivity to extreme values. The mean is calculated by summing all the values in a dataset and then dividing by the number of values. It considers every data point and can be highly influenced by outliers. On the other hand, the median is the middle value in a sorted dataset, separating the higher and lower halves. It is less affected by extreme values and provides a measure of the central tendency that allows for a better representation of the data when outliers are present.

In summary, the median is often used for imputing missing numeric data in data wrangling due to its robustness to outliers and its ability to preserve the general distribution. The choice between mean and median depends on the nature of the data and the objective of the analysis, considering factors such as data distribution and the presence of outliers.

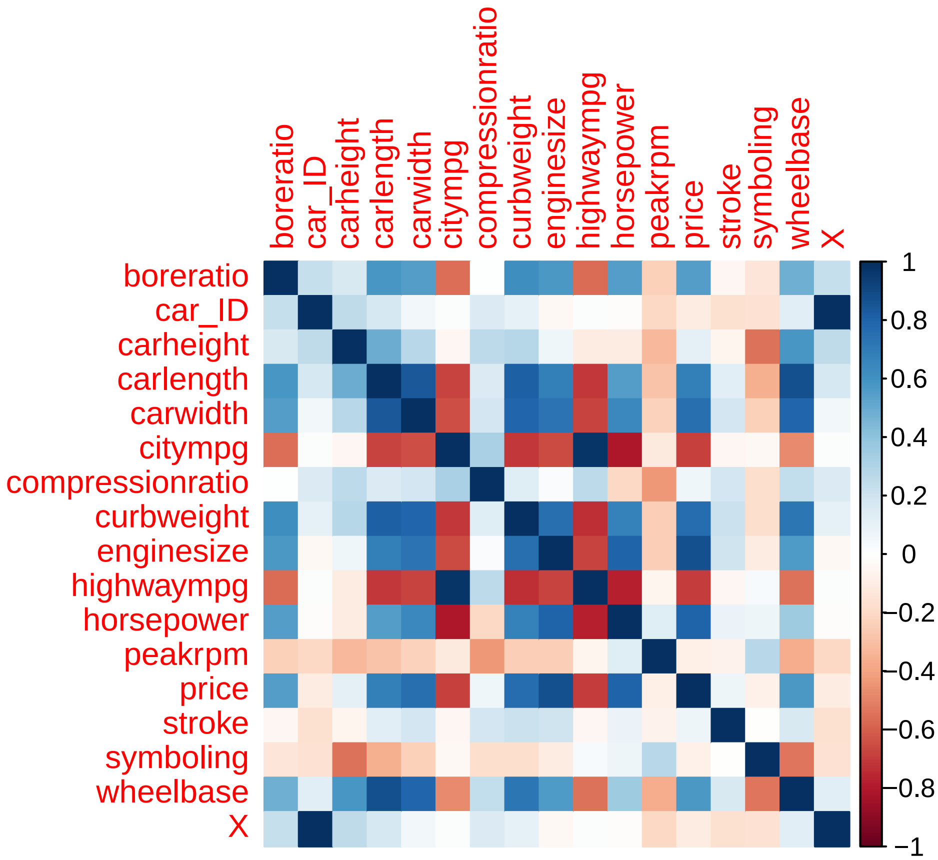
And also, obviously, it is appropriate to use mode to replace missing data in categorical data.

Correlation map

```r
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```r
numeric_data <- my_dataset[,sapply(my_dataset,is.numeric)]
corr_mat <- cor(numeric_data)
corrplot(corr_mat,method = "color",order = 'alphabet')
```



Now we design 4 hypothesis tests and make a decision for it with a confidence level of 5% .

First, we form a multiple linear regression :

```r
mlr_model  <- lm(price ~ symboling
            + fueltype + aspiration + doornumber + carbody + drivewheel + enginelocation + wheelbase + c
            +cylindernumber + enginesize + fuelsystem + boreratio + stroke + compressionratio + horsepow

model_summary <- summary(mlr_model)

# Specify the column names of the features you're interested in
feature_names <- c("stroke", "boreratio", "wheelbase","carheight")

# Extract the t-values of the selected features
t_values <- model_summary$coefficients[feature_names, "t value"]

# Print the t-values
print(t_values)
```

```
##      stroke  boreratio  wheelbase  carheight
## -5.4764120 -3.2809108  0.5973788  1.1171539
```

If variable answer that here $Price$, be a linear function of the said variables , that is, we have a $Multiple\ Linear\ Regression$ (MLR) problem and the relationship between the response variable and predictor variables will be like this :

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n$$

Now we design hypothesis tests.

Hypothesis test 1:

$$H_0 : \beta_{stroke} = 0$$
$$H_1 : \beta_{stroke} \neq 0$$

Thus, the null hypothesis is set that there is no relationship between response and the predictor variables.

According to $t - value$ obtained in the above section , we have:

$$|t_{stroke}| > t_\alpha$$

and we can reject H_0.

Hypothesis test 2:

$$H_0 : \beta_{boreratio} = 0$$
$$H_1 : \beta_{boreratio} \neq 0$$

According to $t - value$ obtained in the above section , we have:

$$|t_{boreratio}| > t_\alpha$$

and we can reject H_0.

Hypothesis test 3:

$$H_0 : \beta_{wheelbase} = 0$$
$$H_1 : \beta_{wheelbase} \neq 0$$

According to $t - value$ obtained in the above section , we have:

$$|t_{wheelbase}| < t_\alpha$$

and we can accept H_0.

Hypothesis test 4:

$$H_0 : \beta_{carheight} = 0$$
$$H_1 : \beta_{carheight} \neq 0$$

According to $t - value$ obtained in the above section , we have:

$$|t_{carheight}| < t_\alpha$$

and we can accept H_0.

$define\ dummy\ variable$

In this section, we define *dummy* variables for categorical columns using the one hot encode method.

```
library('fastDummies')
new_dataset <- dummy_cols(my_dataset, select_columns = c('fueltype','aspiration','doornumber','carbody'
, 'fuelsystem'),remove_selected_columns = TRUE)
```

more about **One hot encoding** :

One-hot encoding is a common method used to convert categorical variables into dummy variables for statistical analysis. It is a technique that creates binary variables representing each category of a categorical variable.

Here's an explanation of the one-hot encoding process:

1. Identify the categorical variable: Determine the categorical variable in your dataset that you want to encode. For example, let's suppose you have a variable called "color" that can have three categories: red, blue, and green.
2. Create dummy variables: Create a new binary variable for each category of the categorical variable. In our example, we would create three dummy variables: "color_red," "color_blue," and "color_green."
3. Assign values: Assign a value of 1 to the dummy variable representing the category to which each observation belongs, and assign a value of 0 to all other dummy variables. For instance, if an observation has a color value of "red," the "color_red" dummy variable would be 1, while the "color_blue" and "color_green" dummy variables would be 0.
4. Interpretation: The resulting binary variables can now be used for statistical analysis, such as regression models. Each dummy variable represents the presence (1) or absence (0) of a specific category, allowing you to examine the effects of each category independently.

One-hot encoding ensures that the categorical variables do not have an inherent order or magnitude assigned to them. Each category is represented by a separate dummy variable, which eliminates any ordinal interpretation.

Remember to handle any reference-level category appropriately. To avoid issues with multicollinearity in regression models, typically, one category is considered the reference level, meaning its corresponding dummy variable is excluded and can be derived from the others.

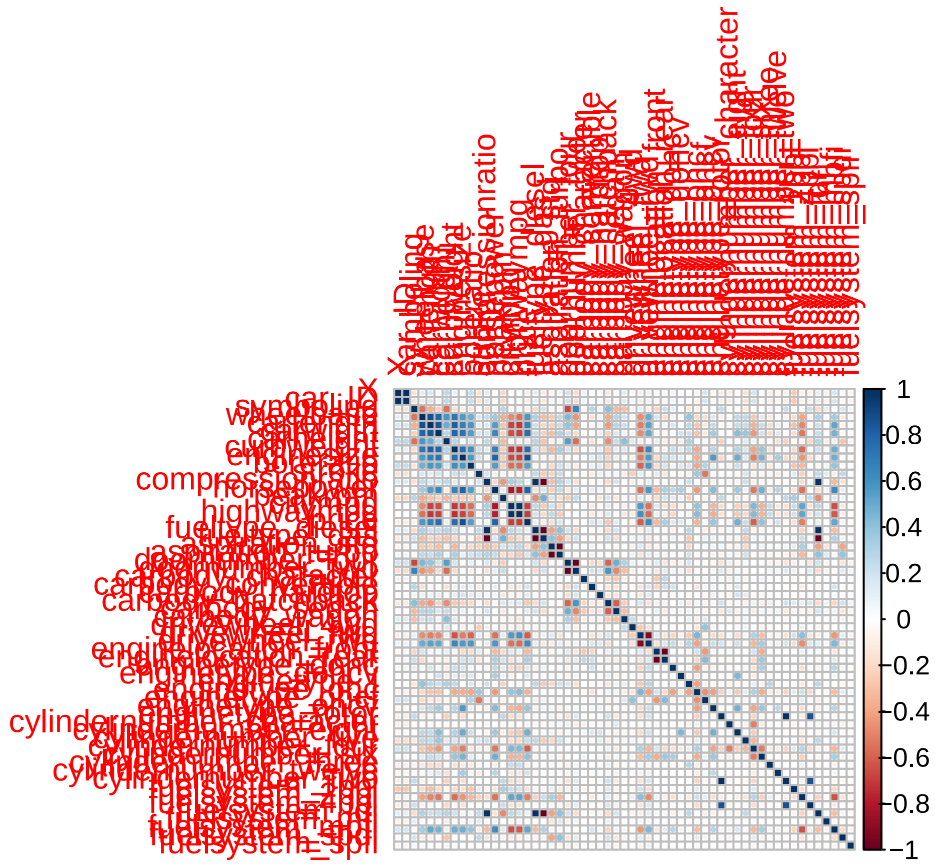
By using one-hot encoding, you can convert categorical variables into a format that can be easily utilized in statistical analyses, allowing for the incorporation of categorical information in predictive models effectively.

Then for the new dataset, we draw its correlation map :

```
new_numeric_data <- new_dataset[,sapply(new_dataset,is.numeric)]

new_corr_mat <- cor(new_numeric_data)

corrplot(new_corr_mat,method = "circle")
```



Dividing the data into two groups, **train** and **test** :

(The ratio of train data to test data is 80 to 20)

```
set.seed(123)  # Setting a seed for reproducibility

# Randomly shuffle the data
shuffled_data <- new_numeric_data[sample(nrow(new_numeric_data)), ]

# Define the proportion of data for training and testing
train_prop <- 0.8  # 80% for training, 20% for test
train_size <- floor(train_prop * nrow(shuffled_data))

# Split the data into training and testing sets
train_data <- shuffled_data[1:train_size, ]
test_data <- shuffled_data[(train_size + 1):nrow(shuffled_data), ]
train_data <- train_data[,-c(0,1,2)]
test_data <- test_data[,-c(0,1,2)]



cat('number of test data = ',nrow(test_data))
```

```
## number of test data =  41
```

```
cat('number of train data = ',nrow(train_data))
```

```
## number of train data =  164
```

Now we have divided the CarPrice_Assignment dataset into a training set and a testing set, and we can build and evaluate our models using separate data subsets.

In this section, we fit the multiple regression model on train data.

```
#selected_columns <- train_data[,4:ncol(new_numeric_data)]
MLR_model <- lm(price ~ . , data = train_data)
Model_summary <- summary(MLR_model)
```

Now,we fit the model on the **train data** ,then calculate and report RSS (Residual Sum of Squares), TSS (Total Sum of Squares), MSE (Mean Squared Error), R-squared, and adjusted R-squared :

```
# Make predictions on train data
predicted <- predict(MLR_model, newdata = train_data)
```

```
## Warning in predict.lm(MLR_model, newdata = train_data): prediction from a
## rank-deficient fit may be misleading
```

```
# Calculate residuals
residuals <- train_data$price - predicted

#  RSS
RSS <- sum(residuals^2)

#  TSS
TSS <- sum((train_data$price - mean(train_data$price))^2)

#  MSE
MSE <- mean(residuals^2)
#  R-squared
R_squared <- 1 - RSS/TSS

#  Adjusted R-squared
num_predictors <- length(coefficients(MLR_model)) - 1
num_obs <- length(train_data$price)
adjusted_R_squared <- 1 - ((1 - R_squared) * (num_obs - 1)) / (num_obs - num_predictors - 1)

cat("For train data :\n\n","RSS:", RSS, "\n","TSS:", TSS, "\n","MSE:", MSE, "\n","R-squared:", R_square
```

```
## For train data :
##
##  RSS: 657947223
##  TSS: 10572191324
##  MSE: 4011873
##  R-squared: 0.9377662
##  Adjusted R-squared: 0.9069348
```

Then,we fit the model on the **test data** and calculate and report RSS (Residual Sum of Squares), TSS (Total Sum of Squares), MSE (Mean Squared Error), R-squared, and adjusted R-squared :

```
# Make predictions on test data
predicted2 <- predict(MLR_model, newdata = test_data)
```

```
## Warning in predict.lm(MLR_model, newdata = test_data): prediction from a
## rank-deficient fit may be misleading
```

```
# Calculate residuals
residuals2 <- test_data$price - predicted2
```

```r
# RSS
RSS <- sum(residuals2^2)

# TSS
TSS <- sum((test_data$price - mean(test_data$price))^2)

# MSE
MSE <- mean(residuals2^2)

# R-squared
R_squared <- 1 - RSS/TSS

# Adjusted R-squared

num_predictors <- length(coefficients(MLR_model)) - 1
num_obs <- length(test_data$price)
adjusted_R_squared <- 1 - ((1 - R_squared) * (num_obs - 1)) / (num_obs - num_predictors - 1)

cat("For test data :\n\n","RSS:", RSS, "\n","TSS:", TSS, "\n","MSE:", MSE, "\n","R-squared:", R_squared
```

```
## For test data :
##
##  RSS: 548477907
##  TSS: 2433455205
##  MSE: 13377510
##  R-squared: 0.7746094
##  Adjusted R-squared: 1.643973
```

An explanation of each of the metrics and their significance in evaluating a regression model:

1. RSS (Residual Sum of Squares):

- RSS represents the sum of the squared differences between the predicted values and the actual values (residuals) in a regression model.
- It measures the overall model's fit, with lower RSS indicating better model performance.
- Helps assess the goodness-of-fit by quantifying the amount of unexplained variability in the dependent variable.

2. TSS (Total Sum of Squares):

- TSS represents the sum of the squared differences between the actual values and the mean of the dependent variable.
- It serves as a benchmark for evaluating the amount of total variability present in the dependent variable.
- Helps calculate the proportion of variability explained by the regression model, known as the R-squared.

3. MSE (Mean Squared Error):

- MSE represents the mean of the squared differences between the predicted values and the actual values.
- It measures the average magnitude of the errors in the model's predictions.
- Similar to RSS, lower MSE values indicate better model performance.

4. R-squared:

- R-squared is a statistical measure that represents the proportion of the total variability in the dependent variable that is explained by the independent variables (predictors).
- It ranges from 0 to 1, where 0 indicates no explanatory power, and 1 represents a perfect fit.
- R-squared provides an indication of how well the model captures the variation in the target variable.

5. Adjusted R-squared:

- Adjusted R-squared is an extension of R-squared that takes into account the number of predictors in the model.
- It penalizes the addition of unnecessary predictors that do not contribute significantly to the model's performance.
- Adjusted R-squared compensates for the potential overfitting caused by including too many predictors and provides a more reliable measure for model comparison.

These metrics are commonly used to evaluate the performance and goodness-of-fit of regression models. They help assess the accuracy of predictions, the extent of explained variability, and the trade-off between model complexity and performance.

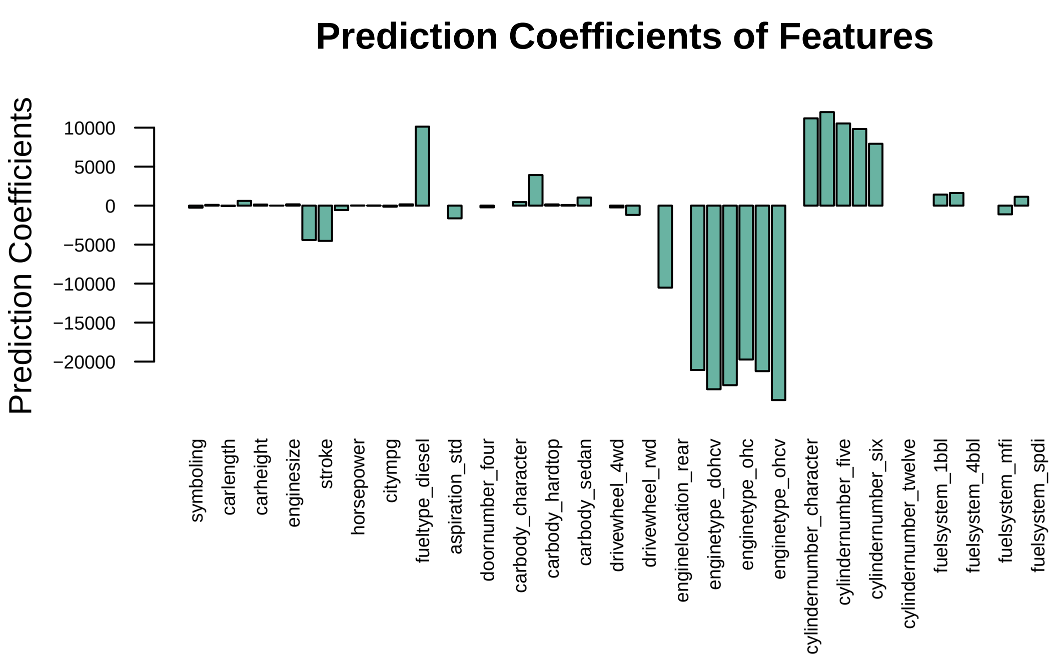
Now, we plot the prediction coefficients of features in our MLR model :

```
coefficients <- coef(MLR_model)

# get feature prediction
feature_coeffs <- coefficients[-1]

# Create the bar plot
par(mar=c(11,4,4,4))

barplot(feature_coeffs, col="#69b3a2",
        names.arg = names(feature_coeffs),   las=2,
        ylab = "Prediction Coefficients",
        main = "Prediction Coefficients of Features",cex.axis=0.6, cex.names=0.6)
```



- In a multiple linear regression model, the magnitude of a coefficient does not necessarily indicate its importance or influence on the target variable (Price). The importance of a coefficient depends on several factors, including the scale of the associated predictor variable and the presence of multicollinearity.

1. Scale of Predictor Variables: When predictor variables are on different scales, the coefficient magnitudes can vary widely. In such cases, comparing the magnitude of coefficients directly may not provide a fair assessment of importance. It is important to normalize or standardize the predictor variables before comparing their coefficients. This ensures that the coefficients are on a comparable scale and allows for a fair evaluation of importance.

2. Multicollinearity: If predictor variables are strongly correlated with each other (multicollinearity), it can affect the interpretation of individual coefficients. In the presence of multicollinearity, the coefficients can become unstable, making it challenging to determine the true importance of each individual variable. In such cases, it may be more appropriate to assess the overall influence of a set of correlated variables rather than focusing solely on individual coefficients.

To determine the importance of predictor variables in a multiple linear regression model, it is advisable to consider other factors such as the statistical significance of coefficients (p-values), the magnitude of standardized coefficients (when variables are standardized), and the overall predictive power of the model (e.g., R-squared, adjusted R-squared).

Additionally, techniques such as feature selection methods (e.g., stepwise regression, LASSO, ridge regression) or domain knowledge can help identify the most important predictors to include in the model.

Therefore, it is essential to interpret the coefficients in the context of the specific variables, their scales, and the presence of multicollinearity, rather than relying solely on their magnitude to determine importance.

- To **improve the performance** of a multiple linear regression model on the CarPrice_Assignment dataset when making predictions on the $test-data$, you can consider the following strategies:

1. Feature Selection: Identify and select the most relevant and informative features for your regression model. You can use techniques like stepwise regression, LASSO (Least Absolute Shrinkage and Selection Operator), or ridge regression to automatically select or penalize less important variables.

2. Address Multicollinearity: Deal with multicollinearity among predictor variables, which can impact model performance and lead to unstable coefficient estimates. You can detect multicollinearity using methods like correlation matrices or variance inflation factor (VIF) analysis and then address it by removing or transforming correlated variables.

3. Data Transformation: Transforming variables can enhance the model's performance. Consider techniques such as log transformations, square root transformations, or scaling to normalize or handle skewed distributions of variables.

4. Outlier Treatment: Identify and handle outliers in the dataset. Outliers can disproportionately influence the regression model's performance. You can remove outliers based on statistical tests or employ robust regression techniques that are less sensitive to outliers.

5. Polynomial or Interaction Terms: Incorporate polynomial features or interaction terms to capture non-linear relationships or interactions between variables if there is evidence of non-linear associations or interaction effects present in the data.

6. Cross-Validation and Regularization: Utilize cross-validation techniques, such as k-fold cross-validation, to assess and tune the model's hyperparameters. Additionally, consider employing regularization techniques like LASSO or ridge regression to shrink or constrain the coefficients, which can help with model generalization and performance on unseen data.

7. Residual Analysis: Analyze the residuals of the model to identify any patterns or deviations from the assumptions of linear regression. Residual plots can provide insights into potential model deficiencies and guide further improvements.
8. Additional Data: Consider gathering additional data that might provide more information relevant to the car price prediction task. Additional features or data points may enhance the model's performance and predictive accuracy.

Remember, iterative model refinement and evaluation with proper validation techniques are essential to find the most effective strategies for improving your specific multiple linear regression model's performance on the CarPrice_Assignment dataset.

Feature Selection

1.Feature Selection using backward method (t-test)

In the backward method , we consider all the features at first, and at each step, we remove the feature that has the largest $p-value$ and continue this process until the $p-value$ of all Variables should be less than 5%.

For this, we take help from the **MASS** library:

```r
library(MASS)
# Perform backward feature selection using t-tests and p-value threshold of 0.05
final_model1 <- stepAIC(MLR_model, direction = "backward", test = "none", alpha = 0.05)
```

```
## Start:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + drivewheel_rwd + enginelocation_front +
##     enginelocation_rear + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     enginetype_rotor + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     cylindernumber_three + cylindernumber_twelve + cylindernumber_two +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_4bbl + fuelsystem_idi +
##     fuelsystem_mfi + fuelsystem_mpfi + fuelsystem_spdi + fuelsystem_spfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + drivewheel_rwd + enginelocation_front +
##     enginelocation_rear + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     enginetype_rotor + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     cylindernumber_three + cylindernumber_twelve + cylindernumber_two +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_4bbl + fuelsystem_idi +
##     fuelsystem_mfi + fuelsystem_mpfi + fuelsystem_spdi
```

```
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + drivewheel_rwd + enginelocation_front +
##     enginelocation_rear + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     enginetype_rotor + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     cylindernumber_three + cylindernumber_twelve + cylindernumber_two +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_4bbl + fuelsystem_idi +
##     fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + drivewheel_rwd + enginelocation_front +
##     enginelocation_rear + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     enginetype_rotor + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     cylindernumber_three + cylindernumber_twelve + cylindernumber_two +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_4bbl + fuelsystem_mfi +
##     fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + drivewheel_rwd + enginelocation_front +
##     enginelocation_rear + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     enginetype_rotor + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     cylindernumber_three + cylindernumber_twelve + cylindernumber_two +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
```

```
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + drivewheel_rwd + enginelocation_front +
##     enginelocation_rear + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     enginetype_rotor + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     cylindernumber_three + cylindernumber_twelve + fuelsystem_1bbl +
##     fuelsystem_2bbl + fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + drivewheel_rwd + enginelocation_front +
##     enginelocation_rear + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     enginetype_rotor + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     cylindernumber_three + fuelsystem_1bbl + fuelsystem_2bbl +
##     fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + drivewheel_rwd + enginelocation_front +
##     enginelocation_rear + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     enginetype_rotor + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + drivewheel_rwd + enginelocation_front +
##     enginelocation_rear + enginetype_dohc + enginetype_dohcv +
```

```
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     cylindernumber_character + cylindernumber_eight + cylindernumber_five +
##     cylindernumber_four + cylindernumber_six + fuelsystem_1bbl +
##     fuelsystem_2bbl + fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + drivewheel_rwd + enginelocation_front +
##     enginetype_dohc + enginetype_dohcv + enginetype_l + enginetype_ohc +
##     enginetype_ohcf + enginetype_ohcv + cylindernumber_character +
##     cylindernumber_eight + cylindernumber_five + cylindernumber_four +
##     cylindernumber_six + fuelsystem_1bbl + fuelsystem_2bbl +
##     fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + carbody_wagon +
##     drivewheel_4wd + drivewheel_fwd + enginelocation_front +
##     enginetype_dohc + enginetype_dohcv + enginetype_l + enginetype_ohc +
##     enginetype_ohcf + enginetype_ohcv + cylindernumber_character +
##     cylindernumber_eight + cylindernumber_five + cylindernumber_four +
##     cylindernumber_six + fuelsystem_1bbl + fuelsystem_2bbl +
##     fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     doornumber_two + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + drivewheel_4wd +
##     drivewheel_fwd + enginelocation_front + enginetype_dohc +
##     enginetype_dohcv + enginetype_l + enginetype_ohc + enginetype_ohcf +
##     enginetype_ohcv + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
```

```
##     fueltype_gas + aspiration_std + aspiration_turbo + doornumber_four +
##     carbody_character + carbody_convertible + carbody_hardtop +
##     carbody_hatchback + carbody_sedan + drivewheel_4wd + drivewheel_fwd +
##     enginelocation_front + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     cylindernumber_character + cylindernumber_eight + cylindernumber_five +
##     cylindernumber_four + cylindernumber_six + fuelsystem_1bbl +
##     fuelsystem_2bbl + fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     fueltype_gas + aspiration_std + doornumber_four + carbody_character +
##     carbody_convertible + carbody_hardtop + carbody_hatchback +
##     carbody_sedan + drivewheel_4wd + drivewheel_fwd + enginelocation_front +
##     enginetype_dohc + enginetype_dohcv + enginetype_l + enginetype_ohc +
##     enginetype_ohcf + enginetype_ohcv + cylindernumber_character +
##     cylindernumber_eight + cylindernumber_five + cylindernumber_four +
##     cylindernumber_six + fuelsystem_1bbl + fuelsystem_2bbl +
##     fuelsystem_mfi + fuelsystem_mpfi
##
##
## Step:  AIC=2575.58
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     aspiration_std + doornumber_four + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_hatchback + carbody_sedan + drivewheel_4wd +
##     drivewheel_fwd + enginelocation_front + enginetype_dohc +
##     enginetype_dohcv + enginetype_l + enginetype_ohc + enginetype_ohcf +
##     enginetype_ohcv + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_mfi + fuelsystem_mpfi
##
##                            Df Sum of Sq       RSS    AIC
## - carbody_hatchback         1     42192 657989415 2573.6
## - carbody_hardtop           1     58406 658005629 2573.6
## - drivewheel_4wd            1    107060 658054283 2573.6
## - doornumber_four           1    563360 658510583 2573.7
## - fuelsystem_mfi            1    994692 658941915 2573.8
## - carbody_character         1   1140774 659087997 2573.9
## - horsepower                1   2602549 660549771 2574.2
## - citympg                   1   2866334 660813557 2574.3
## - fuelsystem_mpfi           1   3064346 661011569 2574.3
## - fuelsystem_1bbl           1   3692662 661639885 2574.5
## - carheight                 1   3814026 661761249 2574.5
## - carlength                 1   4142144 662089367 2574.6
## - wheelbase                 1   4552589 662499812 2574.7
## - compressionratio          1   4666174 662613397 2574.7
## - symboling                 1   4851606 662798829 2574.8
## - highwaympg                1   5141104 663088327 2574.9
## - fuelsystem_2bbl           1   6383206 664330429 2575.2
```

```
## - fueltype_diesel           1   7364181 665311404 2575.4
## <none>                                 657947223 2575.6
## - curbweight                1   8117415 666064638 2575.6
## - drivewheel_fwd            1  10108977 668056200 2576.1
## - carbody_sedan             1  10720397 668667620 2576.2
## - aspiration_std            1  13450134 671397357 2576.9
## - cylindernumber_six        1  14542010 672489233 2577.2
## - cylindernumber_four       1  17980588 675927811 2578.0
## - cylindernumber_five       1  21259287 679206510 2578.8
## - cylindernumber_character  1  24416764 682363987 2579.6
## - carbody_convertible       1  25470817 683418040 2579.8
## - carwidth                  1  28467781 686415004 2580.5
## - cylindernumber_eight      1  33503915 691451138 2581.7
## - peakrpm                   1  39460154 697407377 2583.1
## - enginetype_dohcv          1  42243559 700190782 2583.8
## - boreratio                 1  53375744 711322967 2586.4
## - enginetype_ohc            1  53710913 711658136 2586.4
## - enginetype_ohcf           1  54807645 712754868 2586.7
## - enginetype_dohc           1  58965947 716913170 2587.7
## - enginelocation_front      1  60372726 718319948 2588.0
## - enginetype_l              1  78058089 736005312 2592.0
## - enginetype_ohcv           1  81136361 739083584 2592.7
## - stroke                    1 104739983 762687206 2597.8
## - enginesize                1 237999446 895946669 2624.2
##
## Step:  AIC=2573.59
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     aspiration_std + doornumber_four + carbody_character + carbody_convertible +
##     carbody_hardtop + carbody_sedan + drivewheel_4wd + drivewheel_fwd +
##     enginelocation_front + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     cylindernumber_character + cylindernumber_eight + cylindernumber_five +
##     cylindernumber_four + cylindernumber_six + fuelsystem_1bbl +
##     fuelsystem_2bbl + fuelsystem_mfi + fuelsystem_mpfi
##
##                            Df Sum of Sq       RSS    AIC
## - carbody_hardtop           1     22581 658011996 2571.6
## - drivewheel_4wd            1     96500 658085914 2571.6
## - doornumber_four           1    786844 658776259 2571.8
## - fuelsystem_mfi            1   1007256 658996670 2571.8
## - carbody_character         1   1466927 659456342 2572.0
## - horsepower                1   2694120 660683534 2572.3
## - citympg                   1   2887917 660877331 2572.3
## - fuelsystem_mpfi           1   3091657 661081071 2572.4
## - fuelsystem_1bbl           1   3815396 661804810 2572.5
## - carheight                 1   4095299 662084714 2572.6
## - compressionratio          1   4626643 662616057 2572.7
## - carlength                 1   4763370 662752785 2572.8
## - symboling                 1   4813338 662802753 2572.8
## - wheelbase                 1   4917105 662906519 2572.8
## - highwaympg                1   5112720 663102135 2572.9
## - fuelsystem_2bbl           1   6433894 664423309 2573.2
```

```
## - fueltype_diesel           1   7327587 665317001 2573.4
## <none>                                  657989415 2573.6
## - curbweight                1   8077197 666066612 2573.6
## - drivewheel_fwd            1  10252408 668241822 2574.1
## - aspiration_std            1  13419932 671409347 2574.9
## - cylindernumber_six        1  15520309 673509724 2575.4
## - cylindernumber_four       1  19440412 677429827 2576.4
## - carbody_sedan             1  21268006 679257421 2576.8
## - cylindernumber_five       1  22438495 680427909 2577.1
## - cylindernumber_character  1  26220146 684209560 2578.0
## - carwidth                  1  29287373 687276788 2578.7
## - cylindernumber_eight      1  35785723 693775137 2580.3
## - carbody_convertible       1  36216723 694206138 2580.4
## - peakrpm                   1  41312038 699301452 2581.6
## - enginetype_dohcv          1  44665184 702654599 2582.4
## - boreratio                 1  53398426 711387841 2584.4
## - enginetype_ohc            1  56815391 714804806 2585.2
## - enginetype_ohcf           1  58911089 716900504 2585.7
## - enginelocation_front      1  60557956 718547371 2586.0
## - enginetype_dohc           1  61547848 719537263 2586.3
## - enginetype_l              1  83146260 741135675 2591.1
## - enginetype_ohcv           1  85440868 743430282 2591.6
## - stroke                    1 107448019 765437433 2596.4
## - enginesize                1 237963728 895953143 2622.2
##
## Step:  AIC=2571.6
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     aspiration_std + doornumber_four + carbody_character + carbody_convertible +
##     carbody_sedan + drivewheel_4wd + drivewheel_fwd + enginelocation_front +
##     enginetype_dohc + enginetype_dohcv + enginetype_l + enginetype_ohc +
##     enginetype_ohcf + enginetype_ohcv + cylindernumber_character +
##     cylindernumber_eight + cylindernumber_five + cylindernumber_four +
##     cylindernumber_six + fuelsystem_1bbl + fuelsystem_2bbl +
##     fuelsystem_mfi + fuelsystem_mpfi
##
##                            Df Sum of Sq       RSS    AIC
## - drivewheel_4wd            1     98310 658110306 2569.6
## - doornumber_four           1    820652 658832648 2569.8
## - fuelsystem_mfi            1   1023460 659035456 2569.8
## - carbody_character         1   1444549 659456545 2570.0
## - horsepower                1   2675527 660687523 2570.3
## - citympg                   1   2900882 660912878 2570.3
## - fuelsystem_mpfi           1   3120720 661132716 2570.4
## - fuelsystem_1bbl           1   3809419 661821415 2570.5
## - carheight                 1   4079098 662091094 2570.6
## - compressionratio          1   4633909 662645905 2570.8
## - carlength                 1   4799521 662811517 2570.8
## - symboling                 1   4807069 662819065 2570.8
## - wheelbase                 1   4897027 662909023 2570.8
## - highwaympg                1   5113927 663125923 2570.9
## - fuelsystem_2bbl           1   6449174 664461170 2571.2
## - fueltype_diesel           1   7350829 665362825 2571.4
```

```
## <none>                                658011996 2571.6
## - curbweight                1   8136628 666148624 2571.6
## - drivewheel_fwd            1  10333471 668345467 2572.2
## - aspiration_std            1  13555391 671567387 2572.9
## - cylindernumber_six        1  15800177 673812173 2573.5
## - cylindernumber_four       1  19909886 677921882 2574.5
## - carbody_sedan             1  21421013 679433009 2574.8
## - cylindernumber_five       1  23177772 681189768 2575.3
## - cylindernumber_character  1  26758610 684770606 2576.1
## - carwidth                  1  29290323 687302319 2576.7
## - cylindernumber_eight      1  37076562 695088558 2578.6
## - carbody_convertible       1  37388906 695400902 2578.7
## - peakrpm                   1  41298279 699310275 2579.6
## - enginetype_dohcv          1  45576390 703588386 2580.6
## - boreratio                 1  53430966 711442962 2582.4
## - enginetype_ohc            1  57570184 715582180 2583.3
## - enginetype_ohcf           1  59725844 717737840 2583.8
## - enginelocation_front      1  60564291 718576287 2584.0
## - enginetype_dohc           1  62495578 720507574 2584.5
## - enginetype_l              1  84148459 742160455 2589.3
## - enginetype_ohcv           1  86625827 744637823 2589.9
## - stroke                    1 107441373 765453370 2594.4
## - enginesize                1 242130172 900142168 2621.0
##
## Step:  AIC=2569.62
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     aspiration_std + doornumber_four + carbody_character + carbody_convertible +
##     carbody_sedan + drivewheel_fwd + enginelocation_front + enginetype_dohc +
##     enginetype_dohcv + enginetype_l + enginetype_ohc + enginetype_ohcf +
##     enginetype_ohcv + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_mfi + fuelsystem_mpfi
##
##                            Df Sum of Sq       RSS    AIC
## - doornumber_four           1    980689 659090995 2567.9
## - fuelsystem_mfi            1   1021236 659131541 2567.9
## - carbody_character         1   1461793 659572099 2568.0
## - citympg                   1   3081441 661191747 2568.4
## - fuelsystem_mpfi           1   3174600 661284906 2568.4
## - horsepower                1   3223544 661333849 2568.4
## - fuelsystem_1bbl           1   3903098 662013404 2568.6
## - carheight                 1   4338746 662449051 2568.7
## - compressionratio          1   4623842 662734148 2568.8
## - carlength                 1   4767424 662877729 2568.8
## - wheelbase                 1   5415647 663525953 2569.0
## - highwaympg                1   5462173 663572478 2569.0
## - symboling                 1   5569758 663680064 2569.0
## - fuelsystem_2bbl           1   6509245 664619551 2569.2
## - fueltype_diesel           1   7330498 665440804 2569.4
## - curbweight                1   8045052 666155357 2569.6
## <none>                                658110306 2569.6
## - aspiration_std            1  13457320 671567626 2570.9
```

```
## - drivewheel_fwd           1  13611190 671721496 2571.0
## - cylindernumber_six       1  18524527 676634833 2572.2
## - carbody_sedan            1  21469808 679580114 2572.9
## - cylindernumber_four      1  22740453 680850758 2573.2
## - cylindernumber_five      1  26884794 684995099 2574.2
## - carwidth                 1  29364197 687474502 2574.8
## - cylindernumber_character  1  30303862 688414167 2575.0
## - carbody_convertible      1  38915192 697025497 2577.0
## - peakrpm                  1  41272864 699383170 2577.6
## - cylindernumber_eight     1  44128001 702238306 2578.3
## - boreratio                1  54850766 712961072 2580.8
## - enginetype_dohcv         1  56118994 714229299 2581.0
## - enginelocation_front     1  60531509 718641814 2582.1
## - enginetype_ohc           1  64071262 722181567 2582.9
## - enginetype_ohcf          1  69471846 727582152 2584.1
## - enginetype_dohc          1  69659782 727770088 2584.1
## - enginetype_l             1  89527826 747638131 2588.5
## - enginetype_ohcv          1  96003601 754113907 2589.9
## - stroke                   1 108848608 766958914 2592.7
## - enginesize               1 242068824 900179129 2619.0
##
## Step:  AIC=2567.87
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     aspiration_std + carbody_character + carbody_convertible +
##     carbody_sedan + drivewheel_fwd + enginelocation_front + enginetype_dohc +
##     enginetype_dohcv + enginetype_l + enginetype_ohc + enginetype_ohcf +
##     enginetype_ohcv + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_mfi + fuelsystem_mpfi
##
##                           Df Sum of Sq       RSS    AIC
## - fuelsystem_mfi           1   1086871 660177866 2566.1
## - carbody_character        1   1203421 660294415 2566.2
## - citympg                  1   2788746 661879740 2566.6
## - fuelsystem_mpfi          1   3256536 662347530 2566.7
## - horsepower               1   3734619 662825613 2566.8
## - carheight                1   3895285 662986280 2566.8
## - fuelsystem_1bbl          1   4229927 663320922 2566.9
## - symboling                1   4605806 663696801 2567.0
## - compressionratio         1   4756452 663847446 2567.1
## - highwaympg               1   5117845 664208840 2567.1
## - carlength                1   5365440 664456435 2567.2
## - wheelbase                1   5662030 664753025 2567.3
## - fuelsystem_2bbl          1   6724397 665815392 2567.5
## - curbweight               1   7551904 666642899 2567.7
## - fueltype_diesel          1   7588834 666679829 2567.7
## <none>                               659090995 2567.9
## - aspiration_std           1  13194817 672285812 2569.1
## - drivewheel_fwd           1  15090447 674181442 2569.6
## - cylindernumber_six       1  18615360 677706354 2570.4
## - carbody_sedan            1  20989473 680080467 2571.0
## - cylindernumber_four      1  23007844 682098838 2571.5
```

```
## - cylindernumber_five       1  27267268 686358263 2572.5
## - carwidth                  1  29065969 688156964 2572.9
## - cylindernumber_character  1  30536806 689627801 2573.3
## - carbody_convertible       1  38387040 697478035 2575.2
## - peakrpm                   1  41254086 700345081 2575.8
## - cylindernumber_eight      1  44206522 703297516 2576.5
## - boreratio                 1  54579473 713670468 2578.9
## - enginetype_dohcv          1  56754665 715845660 2579.4
## - enginelocation_front      1  59883362 718974357 2580.1
## - enginetype_ohc            1  64291607 723382602 2581.1
## - enginetype_ohcf           1  70073036 729164031 2582.4
## - enginetype_dohc           1  70098163 729189157 2582.4
## - enginetype_l              1  90177889 749268884 2586.9
## - enginetype_ohcv           1  96606949 755697944 2588.3
## - stroke                    1 110736756 769827751 2591.3
## - enginesize                1 244002841 903093835 2617.5
##
## Step:  AIC=2566.14
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     aspiration_std + carbody_character + carbody_convertible +
##     carbody_sedan + drivewheel_fwd + enginelocation_front + enginetype_dohc +
##     enginetype_dohcv + enginetype_l + enginetype_ohc + enginetype_ohcf +
##     enginetype_ohcv + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_mpfi
##
##                            Df Sum of Sq       RSS    AIC
## - carbody_character         1   1264878 661442744 2564.4
## - citympg                   1   3033346 663211212 2564.9
## - horsepower                1   3771824 663949690 2565.1
## - carheight                 1   3788751 663966617 2565.1
## - fuelsystem_mpfi           1   4581292 664759158 2565.3
## - compressionratio          1   4776289 664954154 2565.3
## - symboling                 1   5037023 665214888 2565.4
## - carlength                 1   5315903 665493769 2565.4
## - fuelsystem_1bbl           1   5469251 665647117 2565.5
## - highwaympg                1   5489477 665667342 2565.5
## - wheelbase                 1   5777993 665955859 2565.6
## - curbweight                1   7587118 667764984 2566.0
## - fueltype_diesel           1   7933103 668110969 2566.1
## <none>                                  660177866 2566.1
## - fuelsystem_2bbl           1   8696771 668874637 2566.3
## - aspiration_std            1  13034615 673212481 2567.3
## - drivewheel_fwd            1  15433974 675611839 2567.9
## - cylindernumber_six        1  18407898 678585763 2568.7
## - carbody_sedan             1  20741600 680919465 2569.2
## - cylindernumber_four       1  22713752 682891618 2569.7
## - cylindernumber_five       1  26989074 687166940 2570.7
## - carwidth                  1  28967159 689145024 2571.2
## - cylindernumber_character  1  30194165 690372031 2571.5
## - carbody_convertible       1  38835465 699013330 2573.5
## - peakrpm                   1  41630694 701808559 2574.2
```

```
## - cylindernumber_eight      1  44174558 704352424 2574.8
## - boreratio                 1  55877880 716055746 2577.5
## - enginetype_dohcv          1  57553391 717731257 2577.8
## - enginelocation_front      1  59757895 719935761 2578.3
## - enginetype_ohc            1  64871235 725049101 2579.5
## - enginetype_ohcf           1  70543423 730721289 2580.8
## - enginetype_dohc           1  70692376 730870242 2580.8
## - enginetype_l              1  90907488 751085354 2585.3
## - enginetype_ohcv           1  97360542 757538408 2586.7
## - stroke                    1 111104787 771282653 2589.7
## - enginesize                1 243113105 903290971 2615.6
##
## Step:  AIC=2564.45
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + citympg + highwaympg + fueltype_diesel +
##     aspiration_std + carbody_convertible + carbody_sedan + drivewheel_fwd +
##     enginelocation_front + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     cylindernumber_character + cylindernumber_eight + cylindernumber_five +
##     cylindernumber_four + cylindernumber_six + fuelsystem_1bbl +
##     fuelsystem_2bbl + fuelsystem_mpfi
##
##                            Df Sum of Sq       RSS    AIC
## - citympg                   1   2723905 664166649 2563.1
## - carheight                 1   3636994 665079738 2563.3
## - horsepower                1   4181995 665624739 2563.5
## - fuelsystem_mpfi           1   4424329 665867072 2563.5
## - carlength                 1   4824324 666267067 2563.6
## - fuelsystem_1bbl           1   4874893 666317637 2563.7
## - compressionratio          1   4961388 666404132 2563.7
## - highwaympg                1   5433392 666876135 2563.8
## - wheelbase                 1   5475416 666918160 2563.8
## - symboling                 1   5936997 667379741 2563.9
## - curbweight                1   7374475 668817219 2564.3
## <none>                                  661442744 2564.4
## - fueltype_diesel           1   8200480 669643223 2564.5
## - fuelsystem_2bbl           1   8366848 669809591 2564.5
## - aspiration_std            1  12509219 673951963 2565.5
## - drivewheel_fwd            1  15894909 677337653 2566.3
## - cylindernumber_six        1  18457900 679900644 2567.0
## - carbody_sedan             1  19559759 681002502 2567.2
## - cylindernumber_four       1  22568749 684011493 2567.9
## - cylindernumber_five       1  26840452 688283196 2569.0
## - carwidth                  1  28604028 690046771 2569.4
## - cylindernumber_character  1  29948027 691390770 2569.7
## - carbody_convertible       1  38168008 699610751 2571.7
## - peakrpm                   1  43986699 705429442 2573.0
## - cylindernumber_eight      1  45205070 706647814 2573.3
## - boreratio                 1  55280830 716723573 2575.6
## - enginetype_dohcv          1  58528919 719971663 2576.4
## - enginetype_ohc            1  64225415 725668158 2577.7
## - enginelocation_front      1  64349942 725792685 2577.7
## - enginetype_ohcf           1  69918836 731361580 2578.9
```

```
## - enginetype_dohc           1  70121101 731563844 2579.0
## - enginetype_l              1  89997692 751440435 2583.4
## - enginetype_ohcv           1  96549065 757991808 2584.8
## - stroke                    1 110497226 771939970 2587.8
## - enginesize                1 241867893 903310636 2613.6
##
## Step:  AIC=2563.13
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + highwaympg + fueltype_diesel + aspiration_std +
##     carbody_convertible + carbody_sedan + drivewheel_fwd + enginelocation_front +
##     enginetype_dohc + enginetype_dohcv + enginetype_l + enginetype_ohc +
##     enginetype_ohcf + enginetype_ohcv + cylindernumber_character +
##     cylindernumber_eight + cylindernumber_five + cylindernumber_four +
##     cylindernumber_six + fuelsystem_1bbl + fuelsystem_2bbl +
##     fuelsystem_mpfi
##
##                            Df Sum of Sq       RSS    AIC
## - highwaympg                1   3416106 667582755 2562.0
## - carheight                 1   3550698 667717347 2562.0
## - carlength                 1   3588507 667755156 2562.0
## - fuelsystem_mpfi           1   4271748 668438397 2562.2
## - symboling                 1   4274614 668441263 2562.2
## - fuelsystem_1bbl           1   4471239 668637888 2562.2
## - horsepower                1   4999977 669166626 2562.4
## - wheelbase                 1   5311797 669478446 2562.4
## - compressionratio          1   6774874 670941523 2562.8
## - curbweight                1   7570515 671737164 2563.0
## - fuelsystem_2bbl           1   7889110 672055758 2563.1
## <none>                                  664166649 2563.1
## - fueltype_diesel           1   9688483 673855132 2563.5
## - aspiration_std            1  11111318 675277967 2563.8
## - drivewheel_fwd            1  15792991 679959640 2565.0
## - cylindernumber_six        1  18383517 682550166 2565.6
## - cylindernumber_four       1  22388371 686555020 2566.6
## - carbody_sedan             1  23222209 687388858 2566.8
## - cylindernumber_five       1  26819408 690986057 2567.6
## - carwidth                  1  28156404 692323053 2567.9
## - cylindernumber_character  1  30009007 694175656 2568.4
## - carbody_convertible       1  39244046 703410695 2570.5
## - peakrpm                   1  42827411 706994060 2571.4
## - cylindernumber_eight      1  42989080 707155729 2571.4
## - boreratio                 1  55029989 719196637 2574.2
## - enginetype_dohcv          1  56007804 720174453 2574.4
## - enginetype_ohc            1  65121336 729287985 2576.5
## - enginelocation_front      1  70638708 734805357 2577.7
## - enginetype_ohcf           1  71536015 735702664 2577.9
## - enginetype_dohc           1  72136203 736302852 2578.0
## - enginetype_l              1  92439015 756605664 2582.5
## - enginetype_ohcv           1  97138653 761305302 2583.5
## - stroke                    1 114916018 779082667 2587.3
## - enginesize                1 239646195 903812844 2611.7
##
## Step:  AIC=2561.97
```

```
## price ~ symboling + wheelbase + carlength + carwidth + carheight +
##     curbweight + enginesize + boreratio + stroke + compressionratio +
##     horsepower + peakrpm + fueltype_diesel + aspiration_std +
##     carbody_convertible + carbody_sedan + drivewheel_fwd + enginelocation_front +
##     enginetype_dohc + enginetype_dohcv + enginetype_l + enginetype_ohc +
##     enginetype_ohcf + enginetype_ohcv + cylindernumber_character +
##     cylindernumber_eight + cylindernumber_five + cylindernumber_four +
##     cylindernumber_six + fuelsystem_1bbl + fuelsystem_2bbl +
##     fuelsystem_mpfi
##
##                            Df Sum of Sq       RSS    AIC
## - carheight                 1   2672028 670254783 2560.6
## - horsepower                1   4266587 671849342 2561.0
## - fuelsystem_mpfi           1   4446771 672029526 2561.1
## - compressionratio          1   4682037 672264792 2561.1
## - wheelbase                 1   4702735 672285490 2561.1
## - fuelsystem_1bbl           1   4793256 672376011 2561.1
## - symboling                 1   5205121 672787876 2561.2
## - carlength                 1   5905292 673488048 2561.4
## - curbweight                1   6557865 674140620 2561.6
## - fueltype_diesel           1   7901316 675484071 2561.9
## <none>                                  667582755 2562.0
## - fuelsystem_2bbl           1   8736357 676319112 2562.1
## - aspiration_std            1  10114386 677697141 2562.4
## - drivewheel_fwd            1  14349941 681932697 2563.4
## - cylindernumber_six        1  21814694 689397449 2565.2
## - carbody_sedan             1  25164109 692746864 2566.0
## - cylindernumber_four       1  26449676 694032431 2566.3
## - cylindernumber_five       1  29275044 696857799 2567.0
## - carwidth                  1  30536558 698119313 2567.3
## - cylindernumber_character  1  34115828 701698583 2568.1
## - carbody_convertible       1  36375846 703958601 2568.7
## - peakrpm                   1  40240741 707823496 2569.6
## - cylindernumber_eight      1  49178088 716760843 2571.6
## - enginetype_dohcv          1  56021330 723604085 2573.2
## - boreratio                 1  56885799 724468554 2573.4
## - enginetype_ohc            1  66139889 733722644 2575.5
## - enginelocation_front      1  71189023 738771778 2576.6
## - enginetype_ohcf           1  73628234 741210989 2577.1
## - enginetype_dohc           1  73724446 741307201 2577.2
## - enginetype_l              1  93824910 761407665 2581.5
## - enginetype_ohcv           1 100126556 767709311 2582.9
## - stroke                    1 114940405 782523161 2586.0
## - enginesize                1 239266546 906849302 2610.2
##
## Step:  AIC=2560.62
## price ~ symboling + wheelbase + carlength + carwidth + curbweight +
##     enginesize + boreratio + stroke + compressionratio + horsepower +
##     peakrpm + fueltype_diesel + aspiration_std + carbody_convertible +
##     carbody_sedan + drivewheel_fwd + enginelocation_front + enginetype_dohc +
##     enginetype_dohcv + enginetype_l + enginetype_ohc + enginetype_ohcf +
##     enginetype_ohcv + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_mpfi
```

```
##
##                            Df Sum of Sq       RSS    AIC
## - carlength                1   3989994 674244777 2559.6
## - horsepower               1   4226329 674481112 2559.7
## - compressionratio         1   5490059 675744842 2560.0
## - symboling                1   5787948 676042731 2560.0
## - curbweight               1   6588177 676842960 2560.2
## - fuelsystem_mpfi          1   7107912 677362695 2560.3
## - wheelbase                1   7127085 677381868 2560.4
## - fuelsystem_1bbl          1   8191668 678446451 2560.6
## <none>                                670254783 2560.6
## - fueltype_diesel          1   9582325 679837108 2560.9
## - aspiration_std           1  10118637 680373420 2561.1
## - fuelsystem_2bbl          1  12076099 682330882 2561.6
## - drivewheel_fwd           1  14093760 684348543 2562.0
## - cylindernumber_six       1  22208569 692463352 2564.0
## - carbody_sedan            1  23667304 693922087 2564.3
## - cylindernumber_four      1  26831545 697086328 2565.1
## - carwidth                 1  28085404 698340187 2565.3
## - cylindernumber_five      1  29753625 700008408 2565.7
## - cylindernumber_character  1  34310112 704564895 2566.8
## - carbody_convertible      1  36119789 706374572 2567.2
## - peakrpm                  1  39174339 709429122 2567.9
## - cylindernumber_eight     1  50432096 720686879 2570.5
## - enginetype_dohcv         1  55282051 725536834 2571.6
## - boreratio                1  57912911 728167694 2572.2
## - enginetype_ohc           1  65918332 736173115 2574.0
## - enginetype_ohcf          1  73565727 743820510 2575.7
## - enginetype_dohc          1  74489713 744744496 2575.9
## - enginelocation_front     1  75738510 745993293 2576.2
## - enginetype_l             1  94320052 764574835 2580.2
## - enginetype_ohcv          1  99344350 769599133 2581.3
## - stroke                   1 116727783 786982566 2584.9
## - enginesize               1 237323747 907578530 2608.3
##
## Step:  AIC=2559.59
## price ~ symboling + wheelbase + carwidth + curbweight + enginesize +
##     boreratio + stroke + compressionratio + horsepower + peakrpm +
##     fueltype_diesel + aspiration_std + carbody_convertible +
##     carbody_sedan + drivewheel_fwd + enginelocation_front + enginetype_dohc +
##     enginetype_dohcv + enginetype_l + enginetype_ohc + enginetype_ohcf +
##     enginetype_ohcv + cylindernumber_character + cylindernumber_eight +
##     cylindernumber_five + cylindernumber_four + cylindernumber_six +
##     fuelsystem_1bbl + fuelsystem_2bbl + fuelsystem_mpfi
##
##                            Df Sum of Sq       RSS    AIC
## - horsepower               1   3116578 677361356 2558.3
## - curbweight               1   3919103 678163881 2558.6
## - wheelbase                1   4192880 678437658 2558.6
## - symboling                1   5439623 679684400 2558.9
## - compressionratio         1   6368176 680612953 2559.1
## - fuelsystem_mpfi          1   7351907 681596684 2559.4
## <none>                                674244777 2559.6
## - fuelsystem_1bbl          1  10527525 684772303 2560.1
```

```
## - fueltype_diesel           1  10532659 684777436 2560.1
## - aspiration_std            1  13290917 687535694 2560.8
## - fuelsystem_2bbl           1  13308757 687553535 2560.8
## - drivewheel_fwd            1  18958643 693203420 2562.1
## - carbody_sedan             1  20785696 695030474 2562.6
## - cylindernumber_six        1  21859393 696104171 2562.8
## - carwidth                  1  24755606 699000383 2563.5
## - cylindernumber_four       1  25398722 699643499 2563.7
## - cylindernumber_five       1  28226150 702470927 2564.3
## - cylindernumber_character  1  32817320 707062097 2565.4
## - carbody_convertible       1  39803785 714048562 2567.0
## - peakrpm                   1  41872392 716117169 2567.5
## - cylindernumber_eight      1  49699952 723944729 2569.3
## - enginetype_dohcv          1  51532736 725777513 2569.7
## - enginetype_ohc            1  64254687 738499464 2572.5
## - boreratio                 1  65887201 740131979 2572.9
## - enginetype_ohcf           1  70865533 745110310 2574.0
## - enginetype_dohc           1  73592530 747837307 2574.6
## - enginelocation_front      1  74030833 748275610 2574.7
## - enginetype_l              1  92311023 766555800 2578.6
## - enginetype_ohcv           1  97574454 771819232 2579.8
## - stroke                    1 112809830 787054607 2583.0
## - enginesize                1 239023858 913268635 2607.4
##
## Step:  AIC=2558.35
## price ~ symboling + wheelbase + carwidth + curbweight + enginesize +
##     boreratio + stroke + compressionratio + peakrpm + fueltype_diesel +
##     aspiration_std + carbody_convertible + carbody_sedan + drivewheel_fwd +
##     enginelocation_front + enginetype_dohc + enginetype_dohcv +
##     enginetype_l + enginetype_ohc + enginetype_ohcf + enginetype_ohcv +
##     cylindernumber_character + cylindernumber_eight + cylindernumber_five +
##     cylindernumber_four + cylindernumber_six + fuelsystem_1bbl +
##     fuelsystem_2bbl + fuelsystem_mpfi
##
##                            Df Sum of Sq       RSS    AIC
## - wheelbase                 1   3266479  680627834 2557.1
## - curbweight                1   4026004  681387360 2557.3
## - symboling                 1   5230138  682591494 2557.6
## <none>                                   677361356 2558.3
## - compressionratio          1   8837217  686198572 2558.5
## - fuelsystem_mpfi           1   8946608  686307963 2558.5
## - fuelsystem_1bbl           1   9731926  687093282 2558.7
## - fueltype_diesel           1  12132781  689494137 2559.3
## - fuelsystem_2bbl           1  12148464  689509819 2559.3
## - cylindernumber_six        1  20494071  697855427 2561.2
## - carbody_sedan             1  20927113  698288469 2561.3
## - drivewheel_fwd            1  21822882  699184237 2561.6
## - cylindernumber_four       1  24364726  701726082 2562.2
## - cylindernumber_five       1  26422587  703783943 2562.6
## - carwidth                  1  29313317  706674672 2563.3
## - cylindernumber_character  1  31572751  708934106 2563.8
## - aspiration_std            1  34239911  711601267 2564.4
## - carbody_convertible       1  37209504  714570860 2565.1
## - cylindernumber_eight      1  47066626  724427982 2567.4
```

```
## - enginetype_dohcv          1  52395184  729756540 2568.6
## - peakrpm                   1  57246946  734608302 2569.7
## - enginetype_ohc            1  66149034  743510389 2571.6
## - boreratio                 1  66231009  743592364 2571.7
## - enginetype_ohcf           1  73537259  750898615 2573.2
## - enginetype_dohc           1  73918333  751279688 2573.3
## - enginetype_ohcv           1  97635174  774996529 2578.4
## - enginetype_l              1  97899018  775260373 2578.5
## - enginelocation_front      1 104292707  781654063 2579.8
## - stroke                    1 125889378  803250733 2584.3
## - enginesize                1 454896083 1132257438 2640.6
##
## Step:  AIC=2557.14
## price ~ symboling + carwidth + curbweight + enginesize + boreratio +
##     stroke + compressionratio + peakrpm + fueltype_diesel + aspiration_std +
##     carbody_convertible + carbody_sedan + drivewheel_fwd + enginelocation_front +
##     enginetype_dohc + enginetype_dohcv + enginetype_l + enginetype_ohc +
##     enginetype_ohcf + enginetype_ohcv + cylindernumber_character +
##     cylindernumber_eight + cylindernumber_five + cylindernumber_four +
##     cylindernumber_six + fuelsystem_1bbl + fuelsystem_2bbl +
##     fuelsystem_mpfi
##
##                            Df Sum of Sq        RSS    AIC
## - curbweight                1   5478545  686106380 2556.4
## - compressionratio          1   7990343  688618177 2557.1
## <none>                                   680627834 2557.1
## - fuelsystem_1bbl           1   9106351  689734185 2557.3
## - fuelsystem_mpfi           1   9773196  690401030 2557.5
## - fueltype_diesel           1  11453685  692081519 2557.9
## - fuelsystem_2bbl           1  12216957  692844791 2558.1
## - symboling                 1  15295919  695923753 2558.8
## - drivewheel_fwd            1  21235377  701863211 2560.2
## - carbody_sedan             1  21501901  702129736 2560.2
## - cylindernumber_six        1  23747234  704375068 2560.8
## - cylindernumber_four       1  26867014  707494848 2561.5
## - cylindernumber_five       1  29267021  709894855 2562.0
## - aspiration_std            1  33527941  714155775 2563.0
## - carbody_convertible       1  33992758  714620592 2563.1
## - cylindernumber_character  1  34642313  715270147 2563.3
## - carwidth                  1  47154492  727782326 2566.1
## - peakrpm                   1  57606520  738234354 2568.5
## - cylindernumber_eight      1  60242246  740870080 2569.1
## - boreratio                 1  63553323  744181157 2569.8
## - enginetype_ohc            1  70976210  751604044 2571.4
## - enginetype_dohcv          1  77231076  757858910 2572.8
## - enginetype_dohc           1  79324821  759952656 2573.2
## - enginetype_ohcf           1  80546808  761174643 2573.5
## - enginetype_l              1  99210195  779838029 2577.5
## - enginelocation_front      1 103040689  783668524 2578.3
## - enginetype_ohcv           1 110962624  791590458 2579.9
## - stroke                    1 133132846  813760680 2584.4
## - enginesize                1 478618891 1159246725 2642.5
##
## Step:  AIC=2556.45
```

```
## price ~ symboling + carwidth + enginesize + boreratio + stroke +
##     compressionratio + peakrpm + fueltype_diesel + aspiration_std +
##     carbody_convertible + carbody_sedan + drivewheel_fwd + enginelocation_front +
##     enginetype_dohc + enginetype_dohcv + enginetype_l + enginetype_ohc +
##     enginetype_ohcf + enginetype_ohcv + cylindernumber_character +
##     cylindernumber_eight + cylindernumber_five + cylindernumber_four +
##     cylindernumber_six + fuelsystem_1bbl + fuelsystem_2bbl +
##     fuelsystem_mpfi
##
##                            Df Sum of Sq        RSS    AIC
## <none>                                   686106380 2556.4
## - compressionratio          1   8926786  695033166 2556.6
## - fuelsystem_1bbl           1   9050788  695157167 2556.6
## - fuelsystem_mpfi           1  10454242  696560622 2556.9
## - fuelsystem_2bbl           1  12283774  698390154 2557.4
## - fueltype_diesel           1  12591439  698697818 2557.4
## - symboling                 1  16933277  703039656 2558.4
## - carbody_sedan             1  21117791  707224171 2559.4
## - drivewheel_fwd            1  27593373  713699753 2560.9
## - aspiration_std            1  37923672  724030052 2563.3
## - carbody_convertible       1  41079502  727185882 2564.0
## - cylindernumber_six        1  43962664  730069044 2564.6
## - cylindernumber_five       1  44673091  730779470 2564.8
## - cylindernumber_four       1  45141368  731247747 2564.9
## - cylindernumber_character  1  55501888  741608268 2567.2
## - peakrpm                   1  61778619  747884998 2568.6
## - carwidth                  1  66901575  753007955 2569.7
## - boreratio                 1  69521540  755627920 2570.3
## - cylindernumber_eight      1  81680134  767786513 2572.9
## - enginelocation_front      1  98539662  784646041 2576.5
## - enginetype_dohcv          1 101584298  787690677 2577.1
## - enginetype_ohc            1 102771430  788877810 2577.3
## - enginetype_ohcf           1 112237327  798343706 2579.3
## - enginetype_dohc           1 114188140  800294520 2579.7
## - stroke                    1 132142671  818249050 2583.3
## - enginetype_l              1 134618437  820724816 2583.8
## - enginetype_ohcv           1 159144913  845251292 2588.7
## - enginesize                1 655219113 1341325492 2664.4
```

```r
pvalus <- summary(final_model1)$coefficients[,"Pr(>|t|)"]
coef <- coef(final_model1)
df1 <- data.frame(Coefficients = names(coef),P_Value = pvalus)
row.names(df1) <-  1:length(coef)
cat(names(coefficients(final_model1)), ':' ,sep = " , ")
```

```
## (Intercept) , symboling , carwidth , enginesize , boreratio , stroke , compressionratio , peakrpm ,
```

```r
cat("Number of selected feautre : ",length(coefficients(final_model1)))
```

```
## Number of selected feautre :  28
```

And as we can see, the number of features has been reduced to 25.

- By reducing the number of features, the complexity of the model will be less and its interpretability will increase. But due to the fact that we have removed some features and ignore their effect, the performance of the model will probably decrease a little.

Now we get the mentioned criteria again in the test and **train data** :

```
# Make predictions on train data
new_predicted1 <- predict(final_model1, newdata = train_data)
residuals <- train_data$price - new_predicted1
RSS <- sum(residuals^2)
TSS <- sum((train_data$price - mean(train_data$price))^2)
MSE <- mean(residuals^2)
R_squared <- 1 - RSS/TSS
num_predictors <- length(coefficients(final_model1)) - 1
num_obs <- length(train_data$price)
adjusted_R_squared <- 1 - ((1 - R_squared) * (num_obs - 1)) / (num_obs - num_predictors - 1)

cat("For train data :\n\n","RSS:", RSS, "\n","TSS:", TSS, "\n","MSE:", MSE, "\n","R-squared:", R_square
```

```
## For train data :
##
##  RSS: 686106380
##  TSS: 10572191324
##  MSE: 4183575
##  R-squared: 0.9351027
##  Adjusted R-squared: 0.9222187
```

and also for **test data** :

```
# Make predictions on test data
new_predicted1 <- predict(final_model1, newdata = test_data)
residuals <- test_data$price - new_predicted1
RSS <- sum(residuals^2)
TSS <- sum((test_data$price - mean(test_data$price))^2)
MSE <- mean(residuals^2)
R_squared <- 1 - RSS/TSS
num_predictors <- length(coefficients(final_model1)) - 1
num_obs <- length(test_data$price)
adjusted_R_squared <- 1 - ((1 - R_squared) * (num_obs - 1)) / (num_obs - num_predictors - 1)

cat("For test data :\n\n","RSS:", RSS, "\n","TSS:", TSS, "\n","MSE:", MSE, "\n","R-squared:", R_squared
```

```
## For test data :
##
##  RSS: 703318146
##  TSS: 2433455205
##  MSE: 17154101
##  R-squared: 0.7109796
##  Adjusted R-squared: 0.1107065
```

As seen in the results above, after performing feature reduction, the following changes are generally expected in the metrics:

1. **RSS** : RSS represents the sum of the squared residuals, which measures the discrepancy between the observed values and the model's predicted values. After feature reduction, if you remove features that were not significantly associated with the response variable, the model's fit will likely improve. This will lead to a increase in the RSS value.
2. **TSS** (Total Sum of Squares): TSS represents the total variation in the response variable. It is the sum of the squared differences between the observed values and the mean of the response variable. Feature reduction does not directly impact TSS, as it measures the total variation in the response variable before and after any modeling.

3. **MSE** : MSE is calculated as the RSS divided by the degrees of freedom. If the feature reduction leads to a increase in the RSS (as discussed above) while maintaining an adequate number of degrees of freedom, the MSE will also increase. An upper MSE indicates that the model's predictions have more error on average.

4. **R-squared**: R-squared represents the proportion of the variance in the response variable explained by the model. It is calculated as 1 minus the ratio of RSS to TSS. After feature reduction, the R-squared value will likely decrease. This indicates that a smaller proportion of the variance in the response is accounted for by the remaining features.

5. **Adjusted R-squared**: Adjusted R-squared adjusts the R-squared value based on the number of predictors in the model. It penalizes the inclusion of additional features that do not contribute significantly to the explanation of variance. After feature reduction, the Adjusted R-squared value may decrease if the removed features were contributing meaningfully to the model's performance.

2.Feature Selection using ANOVA

Now, we perform ANOVA-based feature selection using the F-test and report the 10 most important features :

```
predictors <- subset(train_data, select = -price)
response <- train_data$price

# Calculate the F-scores for each attribute column
f_scores <- sapply(predictors, function(col) {
  summary(aov(response ~ col, data = train_data))[[1]]['F value'][1][,'F value'][1]
})

df2 <- data.frame(Coefficients = names(f_scores),F_Value = f_scores)

# sort rows of df in term of F_Value
sorted_df2 <- df2[order(-df2$F_Value),]
row.names(sorted_df2) <-  1:length(f_scores)

# Print the 10 top attribute and its corresponding F-score
cat("      Feature            F-score\n","1. ", sorted_df2[1,"Coefficients"], "   :   ", sorted_df2[1,"
    ,"2. ", sorted_df2[2,"Coefficients"], "   :   ", sorted_df2[2,"F_Value"], "\n"
,"3. ", sorted_df2[3,"Coefficients"], "   :   ", sorted_df2[3,"F_Value"], "\n",
    "4. ", sorted_df2[4,"Coefficients"], "   :   ", sorted_df2[4,"F_Value"], "\n",
    "5. ", sorted_df2[5,"Coefficients"], "   :   ", sorted_df2[5,"F_Value"], "\n",
    "6. ", sorted_df2[6,"Coefficients"], "   :   ", sorted_df2[6,"F_Value"], "\n",
    "7. ", sorted_df2[7,"Coefficients"], "   :   ", sorted_df2[7,"F_Value"], "\n"
        ,"8. ", sorted_df2[8,"Coefficients"], "   :   ", sorted_df2[8,"F_Value"], "\n"
        ,"9. ", sorted_df2[9,"Coefficients"], "   :   ", sorted_df2[9,"F_Value"], "\n",
    "10. ", sorted_df2[10,"Coefficients"], "   :   ", sorted_df2[10,"F_Value"], "\n")
```

```
##       Feature            F-score
##  1.  enginesize    :    610.2827
##  2.  horsepower    :    302.3904
##  3.  carwidth    :    234.8353
##  4.  curbweight    :    230.1353
##  5.  highwaympg    :    175.863
##  6.  citympg    :    160.1397
##  7.  carlength    :    151.0718
##  8.  drivewheel_rwd    :    109.6815
##  9.  drivewheel_fwd    :    93.27502
##  10.  cylindernumber_four    :    92.76635
```

Synergy Effect

Synergy Effect or Interaction Effect is a phenomenon that arises in the multiple linear regression setting in machine learning, when increase in the value of one Independent variable increases the impact of another Independent variable on the dependent Variable.

for an example in **Advertisement Sales Data-set** :

Let's consider a data-set with advertising budget of a company for different categories and the sales of the company. So, the 3 columns of the data-set are **Radio Advertising**, **Newspaper Advertising** and **Sales**. Let's Fit a linear model in this data-set. This linear model's equation would look like this.

$$Y = \beta_0 + \beta_1 * radio + \beta_2 * newspaper$$

Where beta0 , beta1 and beta2 are the weights and biases to be learnt. Even though this might work fine, There are some cases where this model's predictions might underestimate the actual sales in certain areas and overestimate the actual sales in other areas.

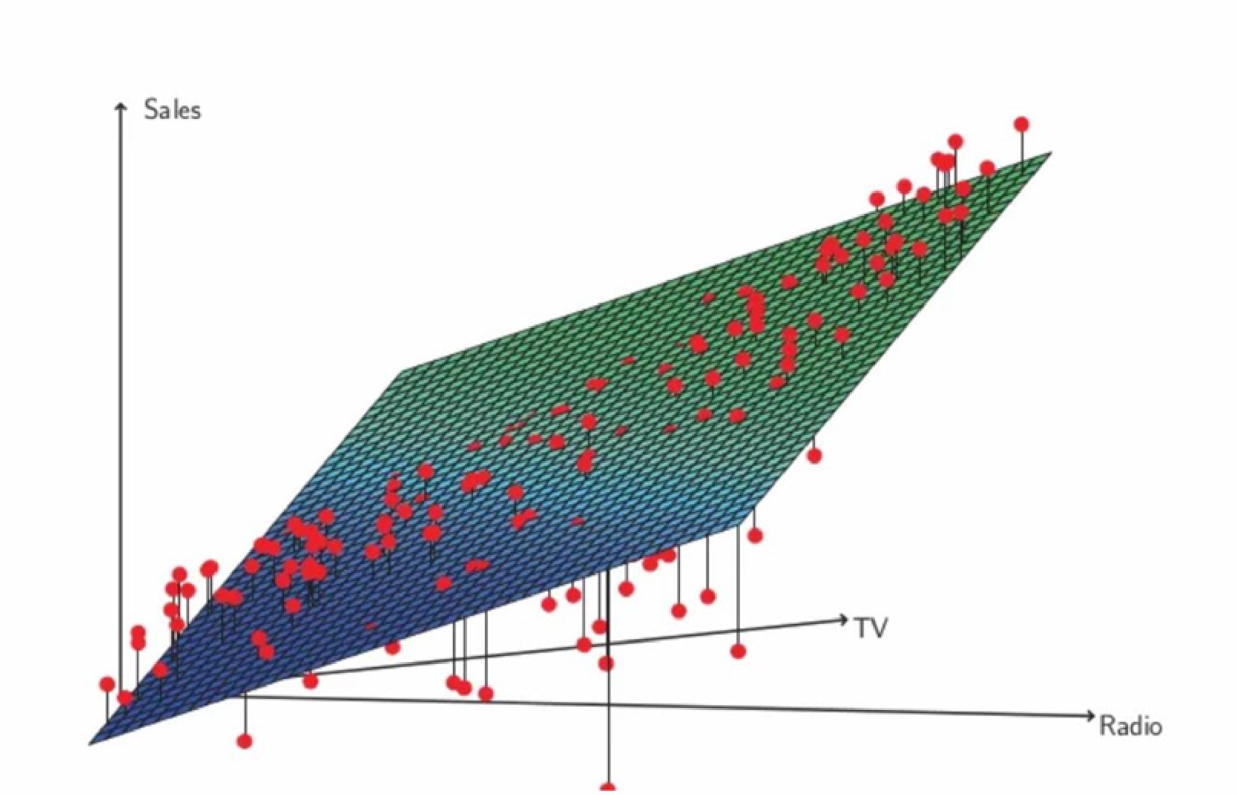


Figure 1: .

In these scenario, the model underestimates sales at times when neither value of radio investment or newspaper investment are extremely small or high.

The model overestimates sales at times when either newspaper or radio has extremely high value i.e, newspaper investment being high and radio investment being low or vice versa. (When one particular predictor value is high).

This is because, Sometimes in reality, the resultant target would have a high value when all the predictors contribute to the prediction instead of just one predictor contributing more and the others contributing less.

CODE:

```
reduced_train_data <- train_data[,c(sorted_df2[1:10,"Coefficients"],"price")]
reduced_test_data <- test_data[,c(sorted_df2[1:10,"Coefficients"],"price")]

library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
feature_combinations <- combn(colnames(reduced_train_data), 2)

#1
new_train_data <- reduced_train_data
new_train_data$newcolumn <- reduced_train_data[,feature_combinations[,1][1]] * reduced_train_data[,featu
model1 <- lm(price ~ . , data = new_train_data)
cat("For synergy effect between ",feature_combinations[,1][1]," and ",feature_combinations[,1][2]," :\n
```

```
## For synergy effect between  enginesize  and  horsepower  :
##  P-Value of new column =  0.07003196
```

becuase :

$$|P - Value| > \frac{\alpha}{2} = 0.025$$

It means that its effect is low and we do not add this variable to the set of features.

```
#2
new_train_data2 <- reduced_train_data
new_train_data2$newcolumn <- reduced_train_data[,feature_combinations[,2][1]] * reduced_train_data[,fea
model2 <- lm(price ~ . , data = new_train_data2)
cat("For synergy effect between ",feature_combinations[,2][1]," and ",feature_combinations[,2][2]," :\n
```

```
## For synergy effect between  enginesize  and  carwidth  :
##  P-Value of new column =  0.002105355
```

becuase :

$$|P - Value| < \frac{\alpha}{2} = 0.025$$

It means that its effect is high and we can add this variable to the set of features.

```
#3
new_train_data3 <- reduced_train_data
new_train_data3$newcolumn <- reduced_train_data[,feature_combinations[,3][1]] * reduced_train_data[,fea
```

```r
model3 <- lm(price ~ . , data = new_train_data3)
cat("For synergy effect between ",feature_combinations[,2][1]," and ",feature_combinations[,2][2]," :\n
```

```
## For synergy effect between  enginesize  and  carwidth  :
##  P-Value of new column =  0.0002901785
```

becuase :

$$|P - Value| < \frac{\alpha}{2} = 0.025$$

It means that its effect is high and we can add this variable to the set of features.

```r
#4
new_train_data4 <- reduced_train_data
new_train_data4$newcolumn <- reduced_train_data[,feature_combinations[,4][1]] * reduced_train_data[,fea
model4 <- lm(price ~ . , data = new_train_data4)
cat("For synergy effect between ",feature_combinations[,4][1]," and ",feature_combinations[,4][2]," :\n
```

```
## For synergy effect between  enginesize  and  highwaympg  :
##  P-Value of new column =  0.0002018706
```

becuase :

$$|P - Value| < \frac{\alpha}{2} = 0.025$$

It means that its effect is high and we can add this variable to the set of features.

```r
#5
new_train_data5 <- reduced_train_data
new_train_data5$newcolumn <- reduced_train_data[,feature_combinations[,5][1]] * reduced_train_data[,fea
model5 <- lm(price ~ . , data = new_train_data5)
cat("For synergy effect between ",feature_combinations[,5][1]," and ",feature_combinations[,5][2]," :\n
```

```
## For synergy effect between  enginesize  and  citympg  :
##  P-Value of new column =  4.163272e-05
```

becuase :

$$|P - Value| < \frac{\alpha}{2} = 0.025$$

It means that its effect is high and we can add this variable to the set of features.

```r
#6
new_train_data6 <- reduced_train_data
new_train_data6$newcolumn <- reduced_train_data[,feature_combinations[,6][1]] * reduced_train_data[,fea
model6 <- lm(price ~ . , data = new_train_data2)
cat("For synergy effect between ",feature_combinations[,6][1]," and ",feature_combinations[,6][2]," :\n
```

```
## For synergy effect between  enginesize  and  carlength  :
##  P-Value of new column =  0.002105355
```

becuase :

$$|P - Value| < \frac{\alpha}{2} = 0.025$$

It means that its effect is high and we can add this variable to the set of features.

```r
#7
new_train_data7 <- reduced_train_data
new_train_data7$newcolumn <- reduced_train_data[,feature_combinations[,7][1]] * reduced_train_data[,fea
model7 <- lm(price ~ . , data = new_train_data7)
cat("For synergy effect between ",feature_combinations[,7][1]," and ",feature_combinations[,7][2]," :\n
```

```
## For synergy effect between  enginesize  and  drivewheel_rwd  :
##  P-Value of new column =  2.979186e-06
```

becuase :

$$|P - Value| < \frac{\alpha}{2} = 0.025$$

It means that its effect is high and we can add this variable to the set of features.

```r
#8
new_train_data8 <- reduced_train_data
new_train_data8$newcolumn <- reduced_train_data[,feature_combinations[,8][1]] * reduced_train_data[,fea
model8 <- lm(price ~ . , data = new_train_data8)
cat("For synergy effect between ",feature_combinations[,8][1]," and ",feature_combinations[,8][2]," :\n
```

```
## For synergy effect between  enginesize  and  drivewheel_fwd  :
##  P-Value of new column =  0.002105355
```

becuase :

$$|P - Value| < \frac{\alpha}{2} = 0.025$$

It means that its effect is high and we can add this variable to the set of features.

```r
#9
new_train_data9 <- reduced_train_data
new_train_data9$newcolumn <- reduced_train_data[,feature_combinations[,9][1]] * reduced_train_data[,fea
model9 <- lm(price ~ . , data = new_train_data9)
cat("For synergy effect between ",feature_combinations[,9][1]," and ",feature_combinations[,9][2]," :\n
```

```
## For synergy effect between  enginesize  and  cylindernumber_four  :
##  P-Value of new column =  1.484377e-06
```

becuase :

$$|P - Value| < \frac{\alpha}{2} = 0.025$$

It means that its effect is high and we can add this variable to the set of features.

```r
#10
new_train_data10 <- reduced_train_data
new_train_data10$newcolumn <- reduced_train_data[,feature_combinations[,10][1]] * reduced_train_data[,f
model10 <- lm(price ~ . , data = new_train_data2)
cat("For synergy effect between ",feature_combinations[,10][1]," and ",feature_combinations[,10][2]," :
```

```
## For synergy effect between  enginesize  and  price  :
##  P-Value of new column =  0.002105355
```

becuase :

$$|P - Value| < \frac{\alpha}{2} = 0.025$$

It means that its effect is high and we can add this variable to the set of features.

Other models :

1.*Decision tree*

about Decision tree :

Decision tree models are popular machine learning algorithms used for both classification and regression tasks. They resemble a flowchart-like structure with a tree-like graph, where each internal node represents a feature, each branch represents a decision rule, and each leaf node represents an outcome or a prediction.

Here's a breakdown of the key concepts and steps in a decision tree model:

1. Feature Selection: The algorithm evaluates different features in your dataset and determines which ones are the most informative for making predictions. It uses various techniques like information gain, Gini impurity, or entropy to measure the predictive power of each feature.
2. Splitting Criteria: The algorithm selects an attribute to split the dataset at each decision node based on a certain criterion. The most commonly used criteria include Gini index, which measures the impurity of a node, and entropy, which quantifies the disorder or uncertainty in the data.
3. Building the Tree: The algorithm recursively splits the data based on the chosen feature and splitting criteria until it reaches the leaf nodes. At each step, it creates child nodes that represent different outcomes or further decisions.
4. Pruning (optional): After building the initial decision tree, pruning techniques can be applied to reduce overfitting. Pruning involves removing unnecessary branches or nodes that are not significantly contributing to the model's predictive power. This helps improve the model's performance on unseen data.
5. Prediction: Once the decision tree is constructed, prediction is performed by traversing the tree from the root node to the appropriate leaf node based on the values of the input features. The outcome associated with the leaf node is then returned as the prediction.

Decision tree models have several advantages, including their interpretability, as the resulting tree structure can be easily visualized and understood. They can handle both numerical and categorical data and can capture non-linear relationships between features. Decision trees are also robust to outliers and missing values.

However, decision tree models can be prone to overfitting, particularly when the tree becomes too complex. Ensemble methods like Random Forests and Gradient Boosting are often used to address this issue by combining multiple decision trees.

CODE:

```r
# Load the required packages
library(rpart)
library(rpart.plot)

# Specify the column names of the predictor variables
predictors <- colnames(train_data)

# Specify the target variable
```

```r
target <- "price"
myformula <- as.formula(paste(target, "~", paste(predictors, collapse = " + ")))
# Train the decision tree model
tree_model <- rpart(myformula, data = train_data)
```

```
## Warning in model.matrix.default(attr(frame, "terms"), frame): the response
## appeared on the right-hand side and was dropped

## Warning in model.matrix.default(attr(frame, "terms"), frame): problem with term
## 15 in model.matrix: no columns are assigned

## Warning in cats * !isord: longer object length is not a multiple of shorter
## object length
```

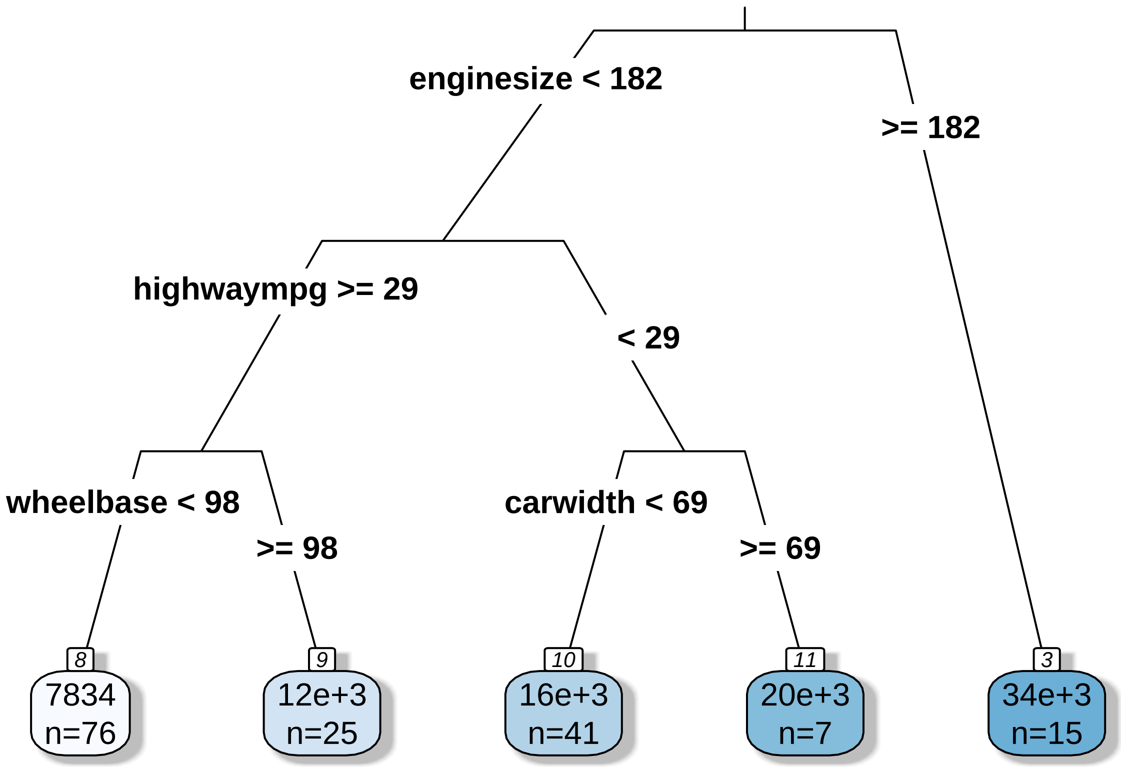
```r
# Predict car prices using the test set
predicted_prices <- predict(tree_model, newdata = test_data)

residuals <- test_data$price - predicted_prices
mse <- mean(residuals^2)
tss <- sum((test_data$price - mean(test_data$price))^2)
rss <- sum(residuals^2)
r_squared <- (tss - rss) / tss

cat("Decision tree model :\n","Mean Squared Error (MSE): ", mse, "\n R-squared: ", r_squared)
```

```
## Decision tree model :
##  Mean Squared Error (MSE):  10949331
##  R-squared:  0.8155205
```

```r
rpart.plot(tree_model,
           box.palette = "Blues",
           nn = TRUE,
           type = 3,
           extra = 1,
           branch = 0.5,
           fallen.leaves = TRUE,
           shadow.col = "gray");
```



$$2.SVM\ (Support\ Vector\ Machines)$$

about SVM :

SVM, short for Support Vector Machines, is a powerful machine learning algorithm commonly used for both classification and regression tasks. It is particularly well-suited for solving complex problems with high-dimensional data. SVM aims to find an optimal hyperplane that separates data points into different classes while maximizing the margin or distance between the classes.

Here are the key concepts and steps involved in SVM:

1. Margin and Hyperplane: The algorithm works by finding a hyperplane that separates the data points of different classes. The hyperplane is essentially a decision boundary in the feature space. The goal is to find the hyperplane that maximizes the margin, which is the distance between the hyperplane and the nearest data points of each class.

2. Support Vectors: SVM gets its name from the data points called support vectors. These are the data points located closest to the decision boundary or the margin. Only these support vectors contribute to the construction of the hyperplane. Other data points in the training set are not used in the decision boundary calculation but may influence it indirectly.

3. Kernel Trick: SVM can handle both linearly separable and non-linearly separable data by utilizing a kernel function. The kernel trick transforms the original feature space into a higher-dimensional space, making it easier to find a hyperplane that can separate the data points. This allows SVM to learn complex decision boundaries.

4. Regularization Parameter (C): C is a hyperparameter in SVM that controls the trade-off between achieving a wider margin and allowing training points to be misclassified. A smaller value of C allows more misclassifications but results in a wider margin, potentially leading to underfitting. On the other hand, a larger C penalizes misclassifications more, resulting in a narrower margin and potential overfitting.
5. Kernel Functions: SVM supports different types of kernel functions such as linear, polynomial, radial basis function (RBF), and sigmoid. These kernel functions define the similarity or distance between data points in the transformed higher-dimensional space. The choice of the kernel function depends on the data and the problem at hand.
6. Prediction: After training, SVM can be used to predict the class labels of new, unseen data points. It assigns a new data point to a particular class based on its position relative to the decision boundary learned during training.

SVM has several advantages, including its ability to handle high-dimensional data, effectiveness in handling non-linear relationships through kernel functions, and robustness against overfitting when the appropriate regularization parameter is chosen.

CODE:

```
library(e1071)
myformula2 <- price ~ wheelbase+carlength+carwidth+carheight+curbweight+enginesize+boreratio+stroke+comp

# Train the SVM model
SVM_model <- svm(myformula2, data = train_data,kernel = "radial")

predicted_prices <- predict(SVM_model, newdata = test_data)
# Calculate the residuals
residuals <- test_data$price - predicted_prices
mse <- mean(residuals^2)
tss <- sum((test_data$price - mean(test_data$price))^2)
rss <- sum((predicted_prices - test_data$price)^2)
r_squared <- 1 - (rss / tss)

cat("SVM model :\n","Mean Squared Error (MSE): ", mse, "\n R-squared: ", r_squared)
```

```
## SVM model :
##  Mean Squared Error (MSE):  6012984
##  R-squared:  0.8986904
```

3.SVR ($Support\ Vector\ Regression$)

Unlike the SVM classifier, SVR is an extension that is specifically designed for regression tasks. It can handle nonlinear relationships and works well in high-dimensional spaces.

```
library(e1071)
# Train the SVR model
SVR_model <- svm(myformula2, data = train_data,kernel = "radial",type = "eps-regression")

predicted_prices <- predict(SVR_model, newdata = test_data)
# Calculate the residuals
residuals <- test_data$price - predicted_prices
mse <- mean(residuals^2)
tss <- sum((test_data$price - mean(test_data$price))^2)
rss <- sum((predicted_prices - test_data$price)^2)
```

```r
r_squared <- 1 - (rss / tss)

cat("SVR model :\n","Mean Squared Error (MSE): ", mse, "\n R-squared: ", r_squared)
```

```
## SVR model :
##  Mean Squared Error (MSE):  6012984
##  R-squared:  0.8986904
```

4.*GBM (Gradient Boosting Machine)*

Gradient Boosting models, such as Gradient Boosting Machines (GBM) or XGBoost, iteratively build an ensemble of weak predictive models. They are known for their strong predictive performance.

```r
library(gbm)
```

```
## Loaded gbm 2.1.8.1
```

```r
# Train the gbm model
gbm_model <- gbm(myformula2, data = train_data,n.trees=1000,interaction.depth=3,shrinkage=0.01)
```

```
## Distribution not specified, assuming gaussian ...
```

```r
predicted_prices <- predict(gbm_model, newdata = test_data,n.trees=1000)
# Calculate the residuals
residuals <- test_data$price - predicted_prices
mse <- mean(residuals^2)
tss <- sum((test_data$price - mean(test_data$price))^2)
rss <- sum((predicted_prices - test_data$price)^2)
r_squared <- 1 - (rss / tss)

cat("GBM model :\n","Mean Squared Error (MSE): ", mse, "\n R-squared: ", r_squared)
```

```
## GBM model :
##  Mean Squared Error (MSE):  5632055
##  R-squared:  0.9051085
```

5.*Random Forest*

Random Forest: Random Forest is an ensemble learning model that combines multiple decision trees to make predictions. It can handle nonlinear relationships and interactions between variables.

```r
library(randomForest)
```

```
## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

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

```r
# Train the rm model
rf_model <- randomForest(myformula2, data = train_data)

predicted_prices <- predict(rf_model, newdata = test_data)
# Calculate the residuals
residuals <- test_data$price - predicted_prices
mse <- mean(residuals^2)
tss <- sum((test_data$price - mean(test_data$price))^2)
rss <- sum((predicted_prices - test_data$price)^2)
r_squared <- 1 - (rss / tss)

cat("RandomForest model :\n","Mean Squared Error (MSE): ", mse, "\n R-squared: ", r_squared)
```

```
## RandomForest model :
##  Mean Squared Error (MSE):  4938574
##  R-squared:  0.9167926
```