Car Analysis

Tarek El-Hajjaoui

2023-02-11

Loading the dataset

```
file_path = '/Users/Tarek/Documents/UCI_MDS_Coding/Stats210P/R_Statistical_Modeling/CarAnalysis/ThreeCa
df = read.table(file_path, header=TRUE, sep="", dec=".")
```

Summary of data set

```
str(df)
```

```
## 'data.frame':    90 obs. of  8 variables:
##  $ CarType: chr  "Porsche" "Porsche" "Porsche" "Porsche" ...
##  $ Price  : num  69.4 56.9 49.9 47.4 42.9 36.9 83 72.9 69.9 67.9 ...
##  $ Age    : int  3 3 2 4 4 6 0 0 2 0 ...
##  $ Mileage: num  21.5 43 19.9 36 44 49.8 1.3 0.67 13.4 9.7 ...
##  $ Car    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Porsche: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Jaguar : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ BMW    : int  0 0 0 0 0 0 0 0 0 0 ...
```

Transforming categorical columns to factor data types.

```
cols <- c("Car", "Porsche", "Jaguar", "BMW")
df[cols] <- lapply(df[cols], as.factor)
```

Checking if column data types transformed successfully.

```
str(df)
```

```
## 'data.frame':    90 obs. of  8 variables:
##  $ CarType: chr  "Porsche" "Porsche" "Porsche" "Porsche" ...
##  $ Price  : num  69.4 56.9 49.9 47.4 42.9 36.9 83 72.9 69.9 67.9 ...
##  $ Age    : int  3 3 2 4 4 6 0 0 2 0 ...
##  $ Mileage: num  21.5 43 19.9 36 44 49.8 1.3 0.67 13.4 9.7 ...
##  $ Car    : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Porsche: Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Jaguar : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ BMW    : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

5 statistic summary

```
summary(df)
```

```
##    CarType              Price            Age             Mileage       Car
##  Length:90          Min.   :12.00   Min.   : 0.000   Min.   :  0.67   0:30
##  Class :character   1st Qu.:23.90   1st Qu.: 3.250   1st Qu.: 20.75   1:30
##  Mode  :character   Median :33.70   Median : 5.000   Median : 42.85   2:30
##                     Mean   :37.58   Mean   : 5.656   Mean   : 41.32
##                     3rd Qu.:49.98   3rd Qu.: 7.000   3rd Qu.: 59.83
##                     Max.   :83.00   Max.   :22.000   Max.   :100.70
##  Porsche Jaguar BMW
##  0:60    0:60   0:60
##  1:30    1:30   1:30
##
##
##
##
```

Scatterplot of Mileage on the X-axis and Price on the Y-axis.

```
ggplot(df, aes(y = Price, x = Mileage, color = CarType)) + geom_point()
```