Automobile

Kaushal Khatiwada

2024-04-19

Assignment 2

```
library("tidyverse")
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library("dplyr")
```

A) Data Collection

Import Retail Store sales dataset in a command-seperated file.

```
car <- read.csv("car dataset.txt", na.strings = "?")
```

B) Data Inspection

```
head(car)
```

```
##   X3  X. alfa.romero gas std  two convertible rwd front X88.60 X168.80 X64.10
## 1  3  NA alfa-romero gas std  two convertible rwd front   88.6   168.8   64.1
## 2  1  NA alfa-romero gas std  two   hatchback rwd front   94.5   171.2   65.5
## 3  2 164        audi gas std four       sedan fwd front   99.8   176.6   66.2
## 4  2 164        audi gas std four       sedan 4wd front   99.4   176.6   66.4
## 5  2  NA        audi gas std  two       sedan fwd front   99.8   177.3   66.3
## 6  1 158        audi gas std four       sedan fwd front  105.8   192.7   71.4
##   X48.80 X2548 dohc four X130 mpfi X3.47 X2.68 X9.00 X111 X5000 X21 X27 X13495
## 1   48.8  2548 dohc four  130 mpfi  3.47  2.68   9.0  111  5000  21  27  16500
```

```
## 2   52.4  2823 ohcv  six  152 mpfi  2.68  3.47   9.0  154  5000  19  26  16500
## 3   54.3  2337  ohc four  109 mpfi  3.19  3.40  10.0  102  5500  24  30  13950
## 4   54.3  2824  ohc five  136 mpfi  3.19  3.40   8.0  115  5500  18  22  17450
## 5   53.1  2507  ohc five  136 mpfi  3.19  3.40   8.5  110  5500  19  25  15250
## 6   55.7  2844  ohc five  136 mpfi  3.19  3.40   8.5  110  5500  19  25  17710
```

```
colnames(car) <- c("symboling", "normalized-losses", "make","fuel-type","aspiration",
                   "num-of-doors","body-style","drive-wheels","engine-location",
                   "wheel-base","length","width","height","curb-weight","engine-type",
                   "num-of-cylinders","engine-size","fuel-system","bore","stroke",
                   "compression-ratio","horsepower","peak-rpm","city-mpg","highway-mpg",
                   "price")
```

1. symboling: -3, -2, -1, 0, 1, 2, 3.
2. normalized-losses: continuous from 65 to 256.
3. make:
   alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercedes-benz, mercury, mitsubishi, nissan, peugot, plymouth, porsche, renault, saab, subaru, toyota, volkswagen, volvo
4. fuel-type: diesel, gas.
5. aspiration: std, turbo.
6. num-of-doors: four, two.
7. body-style: hardtop, wagon, sedan, hatchback, convertible.
8. drive-wheels: 4wd, fwd, rwd.
9. engine-location: front, rear.
10. wheel-base: continuous from 86.6 120.9.
11. length: continuous from 141.1 to 208.1.
12. width: continuous from 60.3 to 72.3.
13. height: continuous from 47.8 to 59.8.
14. curb-weight: continuous from 1488 to 4066.
15. engine-type: dohc, dohcv, l, ohc, ohcf, ohcv, rotor.
16. num-of-cylinders: eight, five, four, six, three, twelve, two.
17. engine-size: continuous from 61 to 326.
18. fuel-system: 1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi.
19. bore: continuous from 2.54 to 3.94.
20. stroke: continuous from 2.07 to 4.17.
21. compression-ratio: continuous from 7 to 23.
22. horsepower: continuous from 48 to 288.
23. peak-rpm: continuous from 4150 to 6600.

24. city-mpg: continuous from 13 to 49.
25. highway-mpg: continuous from 16 to 54.
26. price: continuous from 5118 to 45400.

```
str(car)
```

```
## 'data.frame':    204 obs. of  26 variables:
##  $ symboling        : int  3 1 2 2 2 1 1 1 0 2 ...
##  $ normalized-losses: int  NA NA 164 164 NA 158 NA 158 NA 192 ...
##  $ make             : chr  "alfa-romero" "alfa-romero" "audi" "audi" ...
##  $ fuel-type        : chr  "gas" "gas" "gas" "gas" ...
##  $ aspiration       : chr  "std" "std" "std" "std" ...
##  $ num-of-doors     : chr  "two" "two" "four" "four" ...
##  $ body-style       : chr  "convertible" "hatchback" "sedan" "sedan" ...
##  $ drive-wheels     : chr  "rwd" "rwd" "fwd" "4wd" ...
##  $ engine-location  : chr  "front" "front" "front" "front" ...
##  $ wheel-base       : num  88.6 94.5 99.8 99.4 99.8 ...
##  $ length           : num  169 171 177 177 177 ...
##  $ width            : num  64.1 65.5 66.2 66.4 66.3 71.4 71.4 71.4 67.9 64.8 ...
##  $ height           : num  48.8 52.4 54.3 54.3 53.1 55.7 55.7 55.9 52 54.3 ...
##  $ curb-weight      : int  2548 2823 2337 2824 2507 2844 2954 3086 3053 2395 ...
##  $ engine-type      : chr  "dohc" "ohcv" "ohc" "ohc" ...
##  $ num-of-cylinders : chr  "four" "six" "four" "five" ...
##  $ engine-size      : int  130 152 109 136 136 136 136 131 131 108 ...
##  $ fuel-system      : chr  "mpfi" "mpfi" "mpfi" "mpfi" ...
##  $ bore             : num  3.47 2.68 3.19 3.19 3.19 3.19 3.19 3.13 3.13 3.5 ...
##  $ stroke           : num  2.68 3.47 3.4 3.4 3.4 3.4 3.4 3.4 3.4 2.8 ...
##  $ compression-ratio: num  9 9 10 8 8.5 8.5 8.5 8.3 7 8.8 ...
##  $ horsepower       : int  111 154 102 115 110 110 110 140 160 101 ...
##  $ peak-rpm         : int  5000 5000 5500 5500 5500 5500 5500 5500 5500 5800 ...
##  $ city-mpg         : int  21 19 24 18 19 19 19 17 16 23 ...
##  $ highway-mpg      : int  27 26 30 22 25 25 25 20 22 29 ...
##  $ price            : int  16500 16500 13950 17450 15250 17710 18920 23875 NA 16430 ...
```

Dimensions of dataset.

```
dim(car)
```

```
## [1] 204  26
```

Identifying missing value

```
colSums(is.na(car))
```

```
##         symboling normalized-losses              make         fuel-type
##                 0                40                 0                 0
##        aspiration      num-of-doors        body-style      drive-wheels
##                 0                 2                 0                 0
##   engine-location        wheel-base            length             width
##                 0                 0                 0                 0
```

```
##            height       curb-weight       engine-type  num-of-cylinders
##                 0                 0                 0                 0
##       engine-size       fuel-system              bore            stroke
##                 0                 0                 4                 4
## compression-ratio        horsepower          peak-rpm          city-mpg
##                 0                 2                 2                 0
##       highway-mpg             price
##                 0                 4
```

"normalized-losses": 40 missing data "num-of-doors": 2 missing data "bore": 4 missing data "stroke" : 4 missing data "horsepower": 2 missing data "peak-rpm": 2 missing data "price": 4 missing data

##c) Data Cleaning Handling Missing Values

For Normalized-losses, since it is continuous variable. we will fill NA with its mean value.

```
car %>%
  filter(is.na(`normalized-losses`))
```

```
##    symboling normalized-losses          make fuel-type aspiration num-of-doors
## 1          3                NA   alfa-romero       gas        std          two
## 2          1                NA   alfa-romero       gas        std          two
## 3          2                NA          audi       gas        std          two
## 4          1                NA          audi       gas        std         four
## 5          0                NA          audi       gas      turbo          two
## 6          1                NA           bmw       gas        std         four
## 7          0                NA           bmw       gas        std         four
## 8          0                NA           bmw       gas        std          two
## 9          0                NA           bmw       gas        std         four
## 10         0                NA         isuzu       gas        std         four
## 11         1                NA         isuzu       gas        std          two
## 12         0                NA         isuzu       gas        std         four
## 13         2                NA         isuzu       gas        std          two
## 14         0                NA        jaguar       gas        std         four
## 15         0                NA        jaguar       gas        std          two
## 16         0                NA         mazda    diesel        std         <NA>
## 17         0                NA         mazda    diesel        std         four
## 18        -1                NA mercedes-benz       gas        std         four
## 19         0                NA mercedes-benz       gas        std         four
## 20         1                NA mercedes-benz       gas        std          two
## 21         1                NA       mercury       gas      turbo          two
## 22         3                NA    mitsubishi       gas      turbo          two
## 23         3                NA    mitsubishi       gas      turbo          two
## 24         3                NA    mitsubishi       gas      turbo          two
## 25         0                NA        peugot       gas        std         four
## 26         0                NA        peugot    diesel      turbo         four
## 27         0                NA        peugot       gas        std         four
## 28         0                NA        peugot    diesel      turbo         four
## 29         3                NA      plymouth       gas      turbo          two
## 30         3                NA       porsche       gas        std          two
## 31         3                NA       porsche       gas        std          two
## 32         3                NA       porsche       gas        std          two
## 33         1                NA       porsche       gas        std          two
## 34         0                NA       renault       gas        std         four
```

```
## 35         2                NA       renault       gas        std          two
## 36        -1                NA        toyota       gas        std         four
## 37         3                NA    volkswagen       gas        std          two
## 38         0                NA    volkswagen       gas        std         four
## 39         0                NA    volkswagen    diesel      turbo         four
## 40         0                NA    volkswagen       gas        std         four
##     body-style drive-wheels engine-location wheel-base length width height
## 1  convertible          rwd           front       88.6  168.8  64.1   48.8
## 2    hatchback          rwd           front       94.5  171.2  65.5   52.4
## 3        sedan          fwd           front       99.8  177.3  66.3   53.1
## 4        wagon          fwd           front      105.8  192.7  71.4   55.7
## 5    hatchback          4wd           front       99.5  178.2  67.9   52.0
## 6        sedan          rwd           front      103.5  189.0  66.9   55.7
## 7        sedan          rwd           front      103.5  189.0  66.9   55.7
## 8        sedan          rwd           front      103.5  193.8  67.9   53.7
## 9        sedan          rwd           front      110.0  197.0  70.9   56.3
## 10       sedan          rwd           front       94.3  170.7  61.8   53.5
## 11       sedan          fwd           front       94.5  155.9  63.6   52.0
## 12       sedan          fwd           front       94.5  155.9  63.6   52.0
## 13   hatchback          rwd           front       96.0  172.6  65.2   51.4
## 14       sedan          rwd           front      113.0  199.6  69.6   52.8
## 15       sedan          rwd           front      102.0  191.7  70.6   47.8
## 16       sedan          fwd           front       98.8  177.8  66.5   55.5
## 17       sedan          rwd           front      104.9  175.0  66.1   54.4
## 18       sedan          rwd           front      115.6  202.6  71.7   56.5
## 19       sedan          rwd           front      120.9  208.1  71.7   56.7
## 20     hardtop          rwd           front      112.0  199.2  72.0   55.4
## 21   hatchback          rwd           front      102.7  178.4  68.0   54.8
## 22   hatchback          fwd           front       95.9  173.2  66.3   50.2
## 23   hatchback          fwd           front       95.9  173.2  66.3   50.2
## 24   hatchback          fwd           front       95.9  173.2  66.3   50.2
## 25       wagon          rwd           front      114.2  198.9  68.4   58.7
## 26       wagon          rwd           front      114.2  198.9  68.4   58.7
## 27       wagon          rwd           front      114.2  198.9  68.4   56.7
## 28       wagon          rwd           front      114.2  198.9  68.4   58.7
## 29   hatchback          rwd           front       95.9  173.2  66.3   50.2
## 30     hardtop          rwd            rear       89.5  168.9  65.0   51.6
## 31     hardtop          rwd            rear       89.5  168.9  65.0   51.6
## 32 convertible          rwd            rear       89.5  168.9  65.0   51.6
## 33   hatchback          rwd           front       98.4  175.7  72.3   50.5
## 34       wagon          fwd           front       96.1  181.5  66.5   55.2
## 35   hatchback          fwd           front       96.1  176.8  66.6   50.5
## 36       wagon          rwd           front      104.5  187.8  66.5   54.1
## 37 convertible          fwd           front       94.5  159.3  64.2   55.6
## 38       sedan          fwd           front      100.4  180.2  66.9   55.1
## 39       sedan          fwd           front      100.4  180.2  66.9   55.1
## 40       wagon          fwd           front      100.4  183.1  66.9   55.1
##    curb-weight engine-type num-of-cylinders engine-size fuel-system bore stroke
## 1         2548        dohc             four         130        mpfi 3.47   2.68
## 2         2823        ohcv              six         152        mpfi 2.68   3.47
## 3         2507         ohc             five         136        mpfi 3.19   3.40
## 4         2954         ohc             five         136        mpfi 3.19   3.40
## 5         3053         ohc             five         131        mpfi 3.13   3.40
## 6         3055         ohc              six         164        mpfi 3.31   3.19
```

```
## 7         3230         ohc             six         209        mpfi 3.62   3.39
## 8         3380         ohc             six         209        mpfi 3.62   3.39
## 9         3505         ohc             six         209        mpfi 3.62   3.39
## 10        2337         ohc            four         111        2bbl 3.31   3.23
## 11        1874         ohc            four          90        2bbl 3.03   3.11
## 12        1909         ohc            four          90        2bbl 3.03   3.11
## 13        2734         ohc            four         119        spfi 3.43   3.23
## 14        4066        dohc             six         258        mpfi 3.63   4.17
## 15        3950        ohcv          twelve         326        mpfi 3.54   2.76
## 16        2443         ohc            four         122         idi 3.39   3.39
## 17        2700         ohc            four         134         idi 3.43   3.64
## 18        3740        ohcv           eight         234        mpfi 3.46   3.10
## 19        3900        ohcv           eight         308        mpfi 3.80   3.35
## 20        3715        ohcv           eight         304        mpfi 3.80   3.35
## 21        2910         ohc            four         140        mpfi 3.78   3.12
## 22        2833         ohc            four         156        spdi 3.58   3.86
## 23        2921         ohc            four         156        spdi 3.59   3.86
## 24        2926         ohc            four         156        spdi 3.59   3.86
## 25        3230           l            four         120        mpfi 3.46   3.19
## 26        3430           l            four         152         idi 3.70   3.52
## 27        3285           l            four         120        mpfi 3.46   2.19
## 28        3485           l            four         152         idi 3.70   3.52
## 29        2818         ohc            four         156        spdi 3.59   3.86
## 30        2756        ohcf             six         194        mpfi 3.74   2.90
## 31        2756        ohcf             six         194        mpfi 3.74   2.90
## 32        2800        ohcf             six         194        mpfi 3.74   2.90
## 33        3366       dohcv           eight         203        mpfi 3.94   3.11
## 34        2579         ohc            four         132        mpfi 3.46   3.90
## 35        2460         ohc            four         132        mpfi 3.46   3.90
## 36        3151        dohc             six         161        mpfi 3.27   3.35
## 37        2254         ohc            four         109        mpfi 3.19   3.40
## 38        2661         ohc            five         136        mpfi 3.19   3.40
## 39        2579         ohc            four          97         idi 3.01   3.40
## 40        2563         ohc            four         109        mpfi 3.19   3.40
##    compression-ratio horsepower peak-rpm city-mpg highway-mpg price
## 1                9.0        111     5000       21          27 16500
## 2                9.0        154     5000       19          26 16500
## 3                8.5        110     5500       19          25 15250
## 4                8.5        110     5500       19          25 18920
## 5                7.0        160     5500       16          22    NA
## 6                9.0        121     4250       20          25 24565
## 7                8.0        182     5400       16          22 30760
## 8                8.0        182     5400       16          22 41315
## 9                8.0        182     5400       15          20 36880
## 10               8.5         78     4800       24          29  6785
## 11               9.6         70     5400       38          43    NA
## 12               9.6         70     5400       38          43    NA
## 13               9.2         90     5000       24          29 11048
## 14               8.1        176     4750       15          19 35550
## 15              11.5        262     5000       13          17 36000
## 16              22.7         64     4650       36          42 10795
## 17              22.0         72     4200       31          39 18344
## 18               8.3        155     4750       16          18 34184
## 19               8.0        184     4500       14          16 40960
```

```
## 20                8.0        184     4500       14          16 45400
## 21                8.0        175     5000       19          24 16503
## 22                7.0        145     5000       19          24 12629
## 23                7.0        145     5000       19          24 14869
## 24                7.0        145     5000       19          24 14489
## 25                8.4         97     5000       19          24 12440
## 26               21.0         95     4150       25          25 13860
## 27                8.4         95     5000       19          24 16695
## 28               21.0         95     4150       25          25 17075
## 29                7.0        145     5000       19          24 12764
## 30                9.5        207     5900       17          25 32528
## 31                9.5        207     5900       17          25 34028
## 32                9.5        207     5900       17          25 37028
## 33               10.0        288     5750       17          28    NA
## 34                8.7         NA       NA       23          31  9295
## 35                8.7         NA       NA       23          31  9895
## 36                9.2        156     5200       19          24 15750
## 37                8.5         90     5500       24          29 11595
## 38                8.5        110     5500       19          24 13295
## 39               23.0         68     4500       33          38 13845
## 40                9.0         88     5500       25          31 12290
```

```
car <- car %>%
  mutate(
    `normalized-losses` = replace_na(`normalized-losses`, mean(`normalized-losses`, na.rm = TRUE))
    )
```

For num-of-doors, since it has discrete value. we will fill NA with mode (highest number of occurrences)

```
car %>%
  filter(is.na(`num-of-doors`))
```

```
##   symboling normalized-losses  make fuel-type aspiration num-of-doors
## 1         1               148 dodge       gas      turbo         <NA>
## 2         0               122 mazda    diesel        std         <NA>
##   body-style drive-wheels engine-location wheel-base length width height
## 1      sedan          fwd           front       93.7  157.3  63.8   50.6
## 2      sedan          fwd           front       98.8  177.8  66.5   55.5
##   curb-weight engine-type num-of-cylinders engine-size fuel-system bore stroke
## 1        2191         ohc             four          98        mpfi 3.03   3.39
## 2        2443         ohc             four         122         idi 3.39   3.39
##   compression-ratio horsepower peak-rpm city-mpg highway-mpg price
## 1               7.6        102     5500       24          30  8558
## 2              22.7         64     4650       36          42 10795
```

```
table(car$`num-of-doors`)
```

```
##
## four  two
##  114   88
```

Car with 4 door has highest occurrence.

```
mode <- names(which.max(table(car$`num-of-doors`)))
```

```
car <- car %>%
  mutate(
    `num-of-doors` = ifelse(is.na(`num-of-doors`), mode, `num-of-doors`)
    )
```

For bore, it is also continous variable. we will fill NA with its mean value.

```
car %>%
  filter(is.na(bore))
```

```
##   symboling normalized-losses  make fuel-type aspiration num-of-doors
## 1         3               150 mazda       gas        std          two
## 2         3               150 mazda       gas        std          two
## 3         3               150 mazda       gas        std          two
## 4         3               150 mazda       gas        std          two
##   body-style drive-wheels engine-location wheel-base length width height
## 1  hatchback          rwd           front       95.3    169  65.7   49.6
## 2  hatchback          rwd           front       95.3    169  65.7   49.6
## 3  hatchback          rwd           front       95.3    169  65.7   49.6
## 4  hatchback          rwd           front       95.3    169  65.7   49.6
##   curb-weight engine-type num-of-cylinders engine-size fuel-system bore stroke
## 1        2380       rotor              two          70        4bbl   NA     NA
## 2        2380       rotor              two          70        4bbl   NA     NA
## 3        2385       rotor              two          70        4bbl   NA     NA
## 4        2500       rotor              two          80        mpfi   NA     NA
##   compression-ratio horsepower peak-rpm city-mpg highway-mpg price
## 1               9.4        101     6000       17          23 10945
## 2               9.4        101     6000       17          23 11845
## 3               9.4        101     6000       17          23 13645
## 4               9.4        135     6000       16          23 15645
```

```
car <- car %>%
  mutate(
    bore = replace_na(bore, mean(bore, na.rm = TRUE))
    )
```

For Stoke, it is also continous variable. we will fill NA with its mean value.

```
car %>%
  filter(is.na(stroke))
```

```
##   symboling normalized-losses  make fuel-type aspiration num-of-doors
## 1         3               150 mazda       gas        std          two
## 2         3               150 mazda       gas        std          two
## 3         3               150 mazda       gas        std          two
## 4         3               150 mazda       gas        std          two
##   body-style drive-wheels engine-location wheel-base length width height
## 1  hatchback          rwd           front       95.3    169  65.7   49.6
## 2  hatchback          rwd           front       95.3    169  65.7   49.6
```

```
## 3  hatchback          rwd           front       95.3    169  65.7   49.6
## 4  hatchback          rwd           front       95.3    169  65.7   49.6
##   curb-weight engine-type num-of-cylinders engine-size fuel-system    bore
## 1        2380       rotor              two          70        4bbl 3.32905
## 2        2380       rotor              two          70        4bbl 3.32905
## 3        2385       rotor              two          70        4bbl 3.32905
## 4        2500       rotor              two          80        mpfi 3.32905
##   stroke compression-ratio horsepower peak-rpm city-mpg highway-mpg price
## 1     NA               9.4        101     6000       17          23 10945
## 2     NA               9.4        101     6000       17          23 11845
## 3     NA               9.4        101     6000       17          23 13645
## 4     NA               9.4        135     6000       16          23 15645
```

```
car <- car %>%
  mutate(
    stroke = replace_na(stroke, mean(stroke, na.rm = TRUE))
    )
```

For Horse Power

```
car %>%
  filter(is.na(horsepower))
```

```
##   symboling normalized-losses    make fuel-type aspiration num-of-doors
## 1         0               122 renault       gas        std         four
## 2         2               122 renault       gas        std          two
##   body-style drive-wheels engine-location wheel-base length width height
## 1      wagon          fwd           front       96.1  181.5  66.5   55.2
## 2  hatchback          fwd           front       96.1  176.8  66.6   50.5
##   curb-weight engine-type num-of-cylinders engine-size fuel-system bore stroke
## 1        2579         ohc             four         132        mpfi 3.46    3.9
## 2        2460         ohc             four         132        mpfi 3.46    3.9
##   compression-ratio horsepower peak-rpm city-mpg highway-mpg price
## 1               8.7         NA       NA       23          31  9295
## 2               8.7         NA       NA       23          31  9895
```

```
car <- car %>%
  mutate(
    horsepower = ifelse(is.na(horsepower), mean(horsepower, na.rm = TRUE), horsepower)
  )
```

For Peak RPM

```
car %>%
  filter(is.na(`peak-rpm`))
```

```
##   symboling normalized-losses    make fuel-type aspiration num-of-doors
## 1         0               122 renault       gas        std         four
## 2         2               122 renault       gas        std          two
##   body-style drive-wheels engine-location wheel-base length width height
## 1      wagon          fwd           front       96.1  181.5  66.5   55.2
## 2  hatchback          fwd           front       96.1  176.8  66.6   50.5
```

```
##   curb-weight engine-type num-of-cylinders engine-size fuel-system bore stroke
## 1        2579         ohc             four         132        mpfi 3.46    3.9
## 2        2460         ohc             four         132        mpfi 3.46    3.9
##   compression-ratio horsepower peak-rpm city-mpg highway-mpg price
## 1               8.7   104.2228       NA       23          31  9295
## 2               8.7   104.2228       NA       23          31  9895
```

```
car <- car %>%
  mutate(
    `peak-rpm` = ifelse(is.na(`peak-rpm`), mean(`peak-rpm`, na.rm = TRUE), `peak-rpm`)
  )
```

Price is the output label to predict in basis of these feature to discover pattern. So, row with missing price value must be droped. But still we will fill with the average price.

```
car %>%
  filter(is.na(price))
```

```
##   symboling normalized-losses    make fuel-type aspiration num-of-doors
## 1         0               122    audi       gas      turbo          two
## 2         1               122   isuzu       gas        std          two
## 3         0               122   isuzu       gas        std         four
## 4         1               122 porsche       gas        std          two
##   body-style drive-wheels engine-location wheel-base length width height
## 1  hatchback          4wd           front       99.5  178.2  67.9   52.0
## 2      sedan          fwd           front       94.5  155.9  63.6   52.0
## 3      sedan          fwd           front       94.5  155.9  63.6   52.0
## 4  hatchback          rwd           front       98.4  175.7  72.3   50.5
##   curb-weight engine-type num-of-cylinders engine-size fuel-system bore stroke
## 1        3053         ohc             five         131        mpfi 3.13   3.40
## 2        1874         ohc             four          90        2bbl 3.03   3.11
## 3        1909         ohc             four          90        2bbl 3.03   3.11
## 4        3366       dohcv            eight         203        mpfi 3.94   3.11
##   compression-ratio horsepower peak-rpm city-mpg highway-mpg price
## 1               7.0        160     5500       16          22    NA
## 2               9.6         70     5400       38          43    NA
## 3               9.6         70     5400       38          43    NA
## 4              10.0        288     5750       17          28    NA
```

```
car <- car %>%
  mutate(
    price = ifelse(is.na(price), mean(price, na.rm = TRUE), price)
  )
```

```
colSums(is.na(car))
```

```
##         symboling normalized-losses              make         fuel-type
##                 0                 0                 0                 0
##        aspiration      num-of-doors        body-style      drive-wheels
##                 0                 0                 0                 0
##   engine-location        wheel-base            length             width
##                 0                 0                 0                 0
```

```
##            height       curb-weight       engine-type  num-of-cylinders
##                 0                 0                 0                 0
##       engine-size       fuel-system              bore            stroke
##                 0                 0                 0                 0
## compression-ratio        horsepower          peak-rpm          city-mpg
##                 0                 0                 0                 0
##       highway-mpg             price
##                 0                 0
```

Check Duplicates

```
sum(duplicated(car))
```

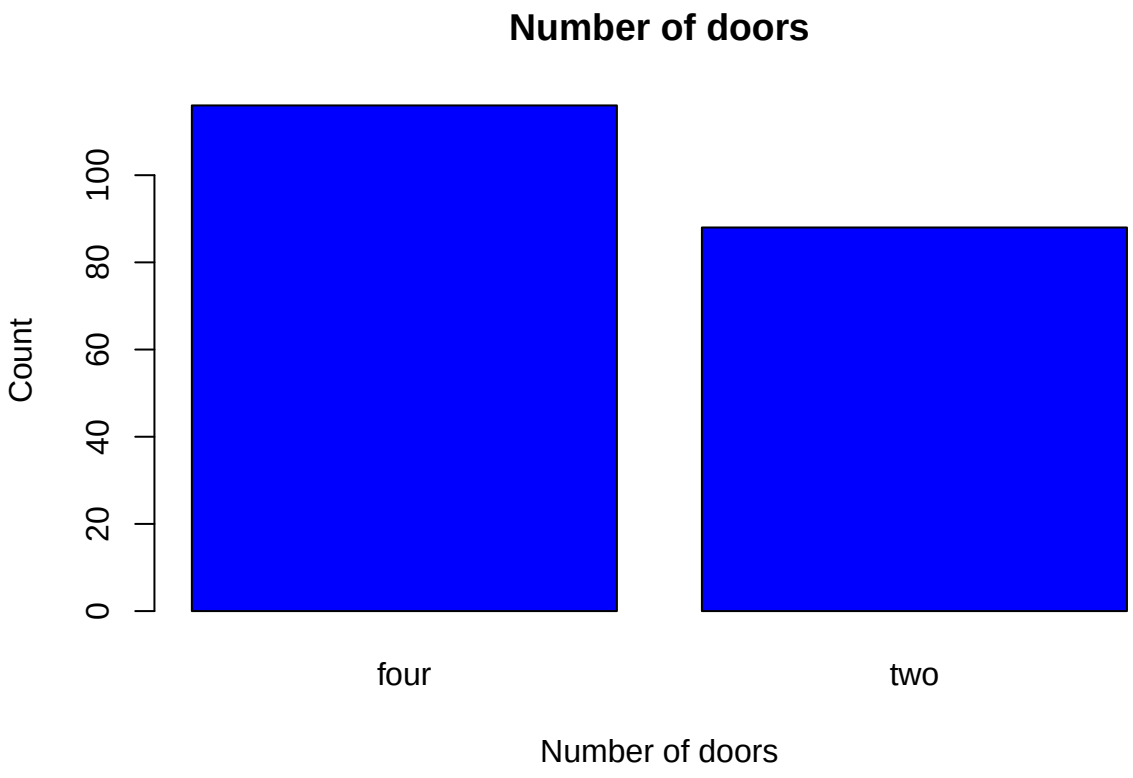
```
## [1] 0
```

EDA

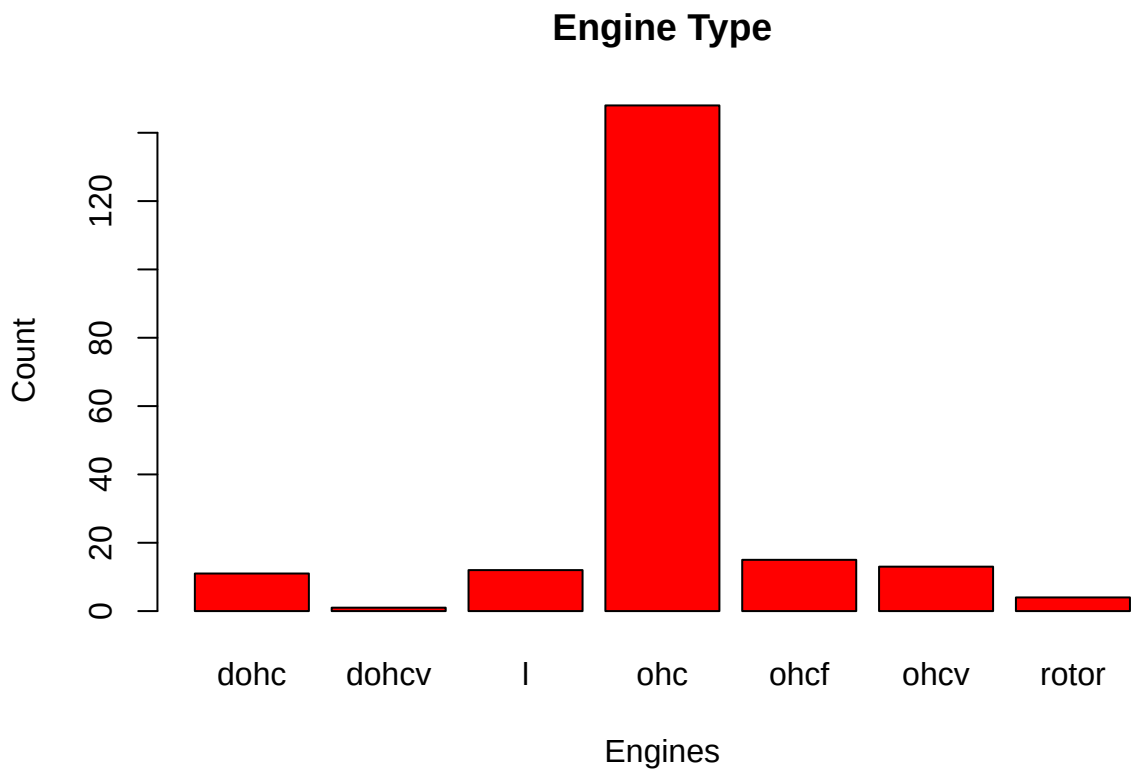
- Univariate Analysis: Studying one variable at a time
- Bivariate Analysis: Studying two variables at a time
- Multivariate Analysis: Studying multiple variables at a time

Univariate Analysis

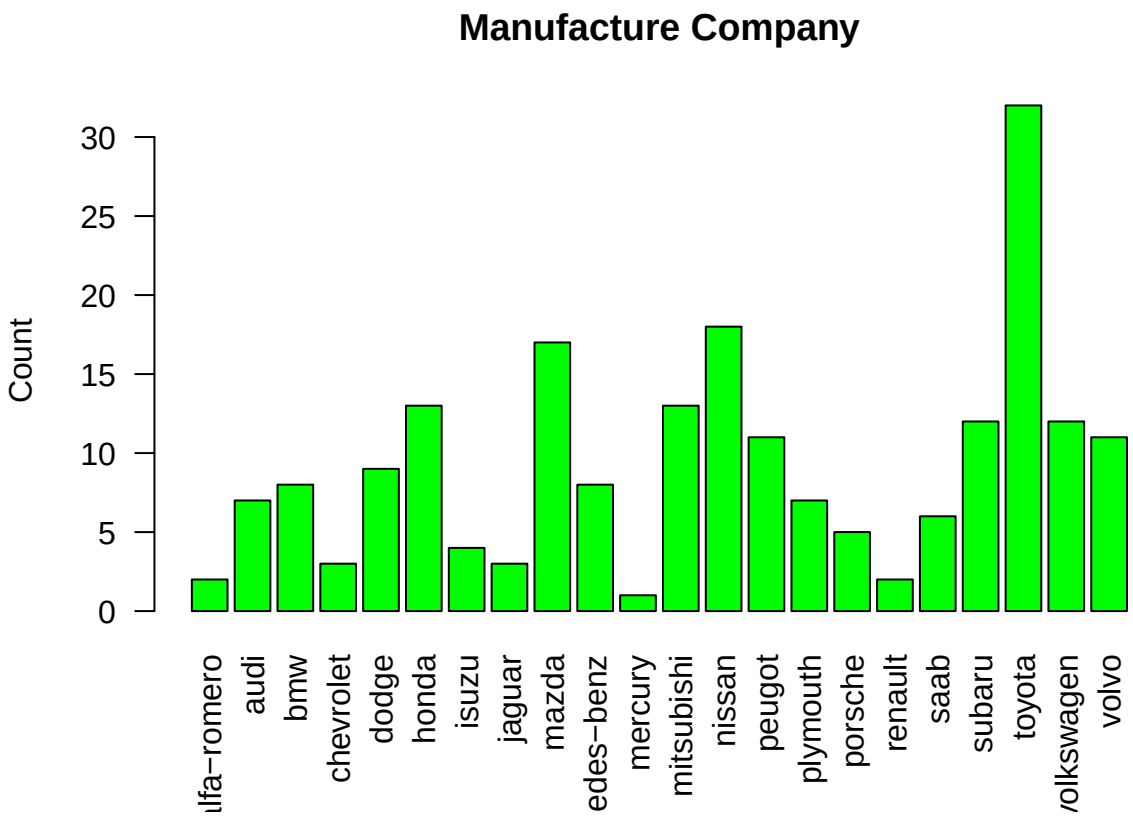
```
door_num <- table(car$`num-of-doors`)
barplot(door_num ,
        main = "Number of doors",
        xlab = "Number of doors",
        ylab = "Count",
        col = "blue")
```




```
engine_type <- table(car$`engine-type`)
barplot(engine_type ,
        main = "Engine Type",
        xlab = "Engines",
        ylab = "Count",
        col = "red")
```

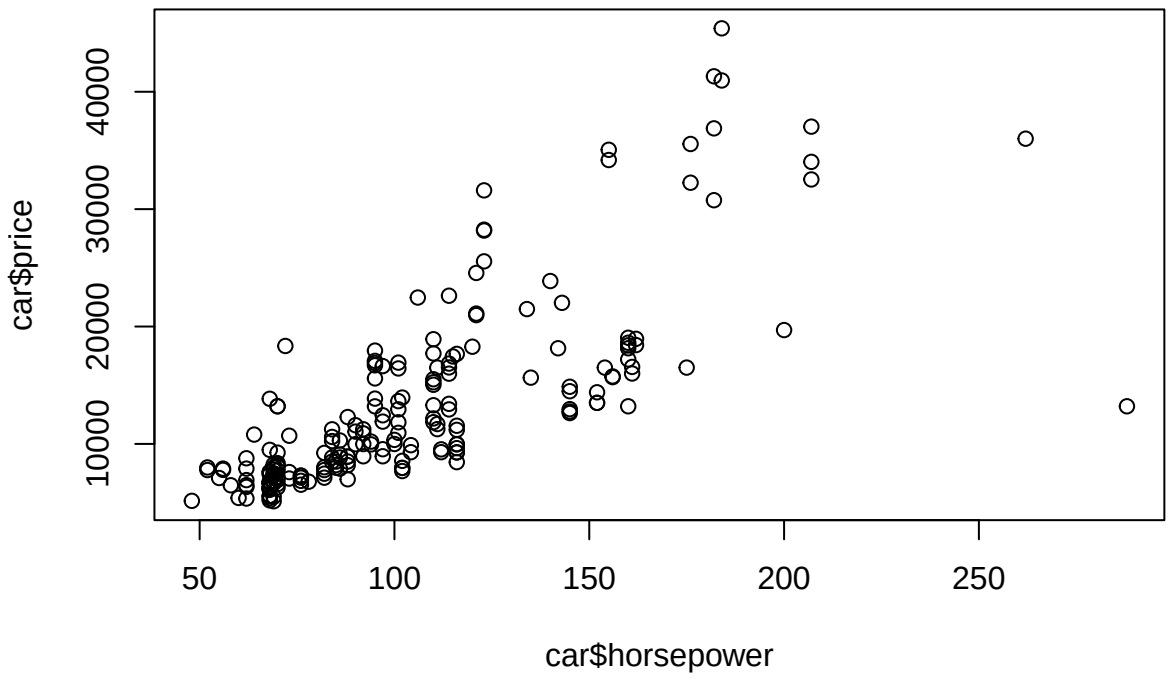



```
makes <- table(car$make)
barplot(makes ,
        main = "Manufacture Company",
        ylab = "Count",
        col = "green",
        las=2)
```



Bivariate Analysis

```
plot(car$horsepower,car$price)
```


car$price
10000
20000
30000
40000
50
100
150
200
250
car$horsepower