Homework 1 - Manay Divatia

2024-02-01

##a

```
cars_df = read.csv("~/Downloads/cars.csv")
```

This dataset contains information about cars. It has whether the car is a sports, suv, wagon, minivan, or pickup car. It also has information on whether it is all wheel drive or rear wheel drive. It also has information on the dealer, engine and other characteristics of the car. It has this information for a variety of cars of different brands. ##b

```
dim(cars_df)
```

```
## [1] 387  18
```

The car has 387 observations (or rows) and 18 variables (or columns). ##c

```
summary(cars_df)
```

```
##      Sports            SUV             Wagon            Minivan
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.1163   Mean   :0.1525   Mean   :0.07235   Mean   :0.05426
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000
##      Pickup       AWD              RWD             Retail
##  Min.   :0   Min.   :0.0000   Min.   :0.0000   Min.   : 10280
##  1st Qu.:0   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 20997
##  Median :0   Median :0.0000   Median :0.0000   Median : 28495
##  Mean   :0   Mean   :0.2016   Mean   :0.2429   Mean   : 33231
##  3rd Qu.:0   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.: 39552
##  Max.   :0   Max.   :1.0000   Max.   :1.0000   Max.   :192465
##      Dealer           Engine        Cylinders        Horsepower
##  Min.   :  9875   Min.   :1.400   Min.   : 3.000   Min.   : 73.0
##  1st Qu.: 19575   1st Qu.:2.300   1st Qu.: 4.000   1st Qu.:165.0
##  Median : 26155   Median :3.000   Median : 6.000   Median :210.0
##  Mean   : 30441   Mean   :3.127   Mean   : 5.757   Mean   :214.4
##  3rd Qu.: 36124   3rd Qu.:3.800   3rd Qu.: 6.000   3rd Qu.:250.0
##  Max.   :173560   Max.   :6.000   Max.   :12.000   Max.   :493.0
##     CityMPG        HighwayMPG        Weight       Wheelbase         Length
##  Min.   :10.00   Min.   :12.00   Min.   :1850   Min.   : 89.0   Min.   :143
##  1st Qu.:18.00   1st Qu.:24.00   1st Qu.:3107   1st Qu.:103.0   1st Qu.:177
##  Median :19.00   Median :27.00   Median :3469   Median :107.0   Median :186
##  Mean   :20.31   Mean   :27.26   Mean   :3532   Mean   :107.2   Mean   :185
##  3rd Qu.:21.50   3rd Qu.:30.00   3rd Qu.:3922   3rd Qu.:112.0   3rd Qu.:193
##  Max.   :60.00   Max.   :66.00   Max.   :6400   Max.   :130.0   Max.   :221
##      Width
##  Min.   :64.00
##  1st Qu.:69.00
```

```
##  Median :71.00
##  Mean   :71.28
##  3rd Qu.:73.00
##  Max.   :81.00
```

From the summary, it looks like all types of cars are there except for the pickup. I know this based on the ranges as it looks like the observation is indicated by a 1 or 0 in the variable of which it is. So for example, a car that is an SUV only would have a 1 in that column and 0s in all other columns for that observation. Additionally, AWD and RWD also have the same 0 or 1 system. There are no null values in the dataset. The horsepower ranges from 73 to 493 and cylinders range from 3 to 12. The weight ranges from 1850 to 6400 and engine from 1.4 to 6. It also looks like the dataframe has information about the miles per gallon which for the city, range from 10 to 60 and for the highway, range from 12 to 66. The weight ranges from 1850 to 6400, wheelbase from 89 to 130, length from 143 to 221, and width from 64 to 81. The retail variable ranges from 10280 to 192465 and dealer from 9875 to 173560 but I think these numbers shouldn't be looked at mathematically because I believe they are more so unique numbers for maybe which dealership the car was bought at. ##d

```r
sum(cars_df$Minivan) / nrow(cars_df)
```

```
## [1] 0.05426357
```

The proportion of cars that were minivans was around 5.4%. I did this by getting the amount of minivans and dividing by the total number of cars which was also the total number of observations. ##e

```r
sum(cars_df$AWD)
```

```
## [1] 78
```

78 cars had all wheel drive. I calculated this by summing up the AWD column because a 1 meant that it was all wheel drive so I could just add up all those. ##f

```r
mean(cars_df$Horsepower)
```

```
## [1] 214.4444
```

The average horsepower is 214.44. I calculated this using the mean function on the Horsepower column. ##g

```r
cars_df$AWDPickup <- ifelse(cars_df$AWD == 1 & cars_df$Pickup == 1, 1, 0)
sum(cars_df$AWDPickup)
```

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

There were no cars that were all wheel drive pickups. I did this by using the ifelse function which has the boolean, and 2 outputs as the parameters. I then saved that vector into a new column I named AWDPickup. I think there are no cars that were all wheel drive pickups because there were no pickups in this dataset at all. So we need more information to know anything about pickups and if they are all wheel drive or not.