Praktik.R

Benhard Leroy

2024-10-04

#Seleksi dan Pembersihan Data  
df <- data.frame(  
 nama = c("Andy","George","Iwan","Jack"),  
 umur = c(20,20,22,21),  
 nilai = c(78,92,85,80)  
)  
df\_selected <- df[, c("nama","nilai")]  
print(df\_selected)

## nama nilai  
## 1 Andy 78  
## 2 George 92  
## 3 Iwan 85  
## 4 Jack 80

df\_filtered <- df[df$umur >20,]  
print(df\_filtered)

## nama umur nilai  
## 3 Iwan 22 85  
## 4 Jack 21 80

df\_sorted <- df[order(-df$nilai), ]  
print(df\_sorted)

## nama umur nilai  
## 2 George 20 92  
## 3 Iwan 22 85  
## 4 Jack 21 80  
## 1 Andy 20 78

#Transformasi Data  
df$nilai\_kategori <- ifelse(df$nilai > 85, "Tinggi","Rendah")  
print(df)

## nama umur nilai nilai\_kategori  
## 1 Andy 20 78 Rendah  
## 2 George 20 92 Tinggi  
## 3 Iwan 22 85 Rendah  
## 4 Jack 21 80 Rendah

library(dplyr)

##   
## Attaching package: 'dplyr'

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

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

df <- mutate(df, nilai\_perc = nilai/100)  
print(df)

## nama umur nilai nilai\_kategori nilai\_perc  
## 1 Andy 20 78 Rendah 0.78  
## 2 George 20 92 Tinggi 0.92  
## 3 Iwan 22 85 Rendah 0.85  
## 4 Jack 21 80 Rendah 0.80

df\_trans <- transmute(df, nama, nilai\_perc)  
print(df\_trans)

## nama nilai\_perc  
## 1 Andy 0.78  
## 2 George 0.92  
## 3 Iwan 0.85  
## 4 Jack 0.80

#Agregasi Data  
df <- data.frame(  
 nama = c("Andy","George","Iwan","Jack"),  
 kelas = c("A","B","A","B"),  
 nilai = c(80,88,90,75)  
)  
df\_summary <- df %>%  
 group\_by(kelas) %>%  
 summarise(mean\_nilai = mean(nilai))  
print(df\_summary)

## # A tibble: 2 × 2  
## kelas mean\_nilai  
## <chr> <dbl>  
## 1 A 85   
## 2 B 81.5

#Handling Missing Value  
df\_kelas <- data.frame(  
 nama = c("Barton","George","Iwan","Jack"),  
 kelas = c("A","B","A","B")  
)  
df\_joined <- left\_join(df, df\_kelas, by = "nama")  
print(df\_joined)

## nama kelas.x nilai kelas.y  
## 1 Andy A 80 <NA>  
## 2 George B 88 B  
## 3 Iwan A 90 A  
## 4 Jack B 75 B

df\_na <- data.frame(  
 nama = c("Barton","George","N/A","Jack"),  
 kelas = c("A","B","A","B")  
)  
df\_clean <- na.omit(df\_na)  
#R Dataset  
mtcars

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160.0 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160.0 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108.0 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258.0 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360.0 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225.0 105 2.76 3.460 20.22 1 0 3 1  
## Duster 360 14.3 8 360.0 245 3.21 3.570 15.84 0 0 3 4  
## Merc 240D 24.4 4 146.7 62 3.69 3.190 20.00 1 0 4 2  
## Merc 230 22.8 4 140.8 95 3.92 3.150 22.90 1 0 4 2  
## Merc 280 19.2 6 167.6 123 3.92 3.440 18.30 1 0 4 4  
## Merc 280C 17.8 6 167.6 123 3.92 3.440 18.90 1 0 4 4  
## Merc 450SE 16.4 8 275.8 180 3.07 4.070 17.40 0 0 3 3  
## Merc 450SL 17.3 8 275.8 180 3.07 3.730 17.60 0 0 3 3  
## Merc 450SLC 15.2 8 275.8 180 3.07 3.780 18.00 0 0 3 3  
## Cadillac Fleetwood 10.4 8 472.0 205 2.93 5.250 17.98 0 0 3 4  
## Lincoln Continental 10.4 8 460.0 215 3.00 5.424 17.82 0 0 3 4  
## Chrysler Imperial 14.7 8 440.0 230 3.23 5.345 17.42 0 0 3 4  
## Fiat 128 32.4 4 78.7 66 4.08 2.200 19.47 1 1 4 1  
## Honda Civic 30.4 4 75.7 52 4.93 1.615 18.52 1 1 4 2  
## Toyota Corolla 33.9 4 71.1 65 4.22 1.835 19.90 1 1 4 1  
## Toyota Corona 21.5 4 120.1 97 3.70 2.465 20.01 1 0 3 1  
## Dodge Challenger 15.5 8 318.0 150 2.76 3.520 16.87 0 0 3 2  
## AMC Javelin 15.2 8 304.0 150 3.15 3.435 17.30 0 0 3 2  
## Camaro Z28 13.3 8 350.0 245 3.73 3.840 15.41 0 0 3 4  
## Pontiac Firebird 19.2 8 400.0 175 3.08 3.845 17.05 0 0 3 2  
## Fiat X1-9 27.3 4 79.0 66 4.08 1.935 18.90 1 1 4 1  
## Porsche 914-2 26.0 4 120.3 91 4.43 2.140 16.70 0 1 5 2  
## Lotus Europa 30.4 4 95.1 113 3.77 1.513 16.90 1 1 5 2  
## Ford Pantera L 15.8 8 351.0 264 4.22 3.170 14.50 0 1 5 4  
## Ferrari Dino 19.7 6 145.0 175 3.62 2.770 15.50 0 1 5 6  
## Maserati Bora 15.0 8 301.0 335 3.54 3.570 14.60 0 1 5 8  
## Volvo 142E 21.4 4 121.0 109 4.11 2.780 18.60 1 1 4 2

?mtcars

## starting httpd help server ...

## done

Data\_Cars <- mtcars  
dim(Data\_Cars)

## [1] 32 11

names(Data\_Cars)

## [1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear"  
## [11] "carb"

rownames(Data\_Cars)

## [1] "Mazda RX4" "Mazda RX4 Wag" "Datsun 710"   
## [4] "Hornet 4 Drive" "Hornet Sportabout" "Valiant"   
## [7] "Duster 360" "Merc 240D" "Merc 230"   
## [10] "Merc 280" "Merc 280C" "Merc 450SE"   
## [13] "Merc 450SL" "Merc 450SLC" "Cadillac Fleetwood"   
## [16] "Lincoln Continental" "Chrysler Imperial" "Fiat 128"   
## [19] "Honda Civic" "Toyota Corolla" "Toyota Corona"   
## [22] "Dodge Challenger" "AMC Javelin" "Camaro Z28"   
## [25] "Pontiac Firebird" "Fiat X1-9" "Porsche 914-2"   
## [28] "Lotus Europa" "Ford Pantera L" "Ferrari Dino"   
## [31] "Maserati Bora" "Volvo 142E"

Data\_Cars$carb

## [1] 4 4 1 1 2 1 4 2 2 4 4 3 3 3 4 4 4 1 2 1 1 2 2 4 2 1 2 2 4 6 8 2

sort(Data\_Cars$carb)

## [1] 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 4 4 4 4 4 4 4 4 4 4 6 8

summary(Data\_Cars)

## mpg cyl disp hp   
## Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0   
## 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5   
## Median :19.20 Median :6.000 Median :196.3 Median :123.0   
## Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7   
## 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0   
## Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0   
## drat wt qsec vs   
## Min. :2.760 Min. :1.513 Min. :14.50 Min. :0.0000   
## 1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000   
## Median :3.695 Median :3.325 Median :17.71 Median :0.0000   
## Mean :3.597 Mean :3.217 Mean :17.85 Mean :0.4375   
## 3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000   
## Max. :4.930 Max. :5.424 Max. :22.90 Max. :1.0000   
## am gear carb   
## Min. :0.0000 Min. :3.000 Min. :1.000   
## 1st Qu.:0.0000 1st Qu.:3.000 1st Qu.:2.000   
## Median :0.0000 Median :4.000 Median :2.000   
## Mean :0.4062 Mean :3.688 Mean :2.812   
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :1.0000 Max. :5.000 Max. :8.000

#Max dan Min  
max(Data\_Cars$carb)

## [1] 8

min(Data\_Cars$carb)

## [1] 1

which.max(Data\_Cars$carb)

## [1] 31

which.min(Data\_Cars$carb)

## [1] 3

rownames(Data\_Cars)[which.max(Data\_Cars$carb)]

## [1] "Maserati Bora"

rownames(Data\_Cars)[which.min(Data\_Cars$carb)]

## [1] "Datsun 710"

data1 <- c(64,25,72,35,64,13,54,46,31,82,64,23,46,47)  
  
#Mean, Median, Mode  
mean(Data\_Cars$mpg)

## [1] 20.09062

median(Data\_Cars$mpg)

## [1] 19.2

names(sort(-table(Data\_Cars$wt)))[1]

## [1] "3.44"

hist(data1,  
 main= "Histogram Sebaran Data",  
 xlab = "Nilai",  
 ylab = "Frekuensi",  
 col="blue",  
 border="red")

