**Assignment Exercise 1 with mtcars dataset (package – “dataset”)**

**DPA Batch II (IMI Delhi), Maximum Marks-15, Last Date of Submission – July 4, 2019**

1. Examine the dataset **mtcars**, and report the data type of each variable in the file. What is the data type of **mtcars** file? How many rows, how many columns are there?

dim(mtcars)

[1] 32 11

> class(mtcars)

[1] "data.frame"

> str(mtcars)

'data.frame': 32 obs. of 11 variables:

$ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...

$ cyl : num 6 6 4 6 8 6 8 4 4 6 ...

$ disp: num 160 160 108 258 360 ...

$ hp : num 110 110 93 110 175 105 245 62 95 123 ...

$ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...

$ wt : num 2.62 2.88 2.32 3.21 3.44 ...

$ qsec: num 16.5 17 18.6 19.4 17 ...

$ vs : num 0 0 1 1 0 1 0 1 1 1 ...

$ am : num 1 1 1 0 0 0 0 0 0 0 ...

$ gear: num 4 4 4 3 3 3 3 4 4 4 ...

$ carb: num 4 4 1 1 2 1 4 2 2 4 ...

*32 rows and 11 columns*

1. What are the row names and column names for the **mtcars** file? Compare your results with row names and column names for **mpg** file. What do you observe?

row.names(mtcars)

[1] "Mazda RX4" "Mazda RX4 Wag"

[3] "Datsun 710" "Hornet 4 Drive"

[5] "Hornet Sportabout" "Valiant"

[7] "Duster 360" "Merc 240D"

[9] "Merc 230" "Merc 280"

[11] "Merc 280C" "Merc 450SE"

[13] "Merc 450SL" "Merc 450SLC"

[15] "Cadillac Fleetwood" "Lincoln Continental"

[17] "Chrysler Imperial" "Fiat 128"

[19] "Honda Civic" "Toyota Corolla"

[21] "Toyota Corona" "Dodge Challenger"

[23] "AMC Javelin" "Camaro Z28"

[25] "Pontiac Firebird" "Fiat X1-9"

[27] "Porsche 914-2" "Lotus Europa"

[29] "Ford Pantera L" "Ferrari Dino"

[31] "Maserati Bora" "Volvo 142E"

colnames(mtcars)

[1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am"

[10] "gear" "carb"

colnames(mpg)

[1] "manufacturer" "model" "displ" "year"

[5] "cyl" "trans" "drv" "cty"

[9] "hwy" "fl" "class"

row.names(mpg)

[1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10"

[11] "11" "12" "13" "14" "15" "16" "17" "18" "19" "20"

[21] "21" "22" "23" "24" "25" "26" "27" "28" "29" "30"

[31] "31" "32" "33" "34" "35" "36" "37" "38" "39" "40"

[41] "41" "42" "43" "44" "45" "46" "47" "48" "49" "50"

[51] "51" "52" "53" "54" "55" "56" "57" "58" "59" "60"

[61] "61" "62" "63" "64" "65" "66" "67" "68" "69" "70"

[71] "71" "72" "73" "74" "75" "76" "77" "78" "79" "80"

[81] "81" "82" "83" "84" "85" "86" "87" "88" "89" "90"

[91] "91" "92" "93" "94" "95" "96" "97" "98" "99" "100"

[101] "101" "102" "103" "104" "105" "106" "107" "108" "109" "110"

[111] "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"

[121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130"

[131] "131" "132" "133" "134" "135" "136" "137" "138" "139" "140"

[141] "141" "142" "143" "144" "145" "146" "147" "148" "149" "150"

[151] "151" "152" "153" "154" "155" "156" "157" "158" "159" "160"

[161] "161" "162" "163" "164" "165" "166" "167" "168" "169" "170"

[171] "171" "172" "173" "174" "175" "176" "177" "178" "179" "180"

[181] "181" "182" "183" "184" "185" "186" "187" "188" "189" "190"

[191] "191" "192" "193" "194" "195" "196" "197" "198" "199" "200"

[201] "201" "202" "203" "204" "205" "206" "207" "208" "209" "210"

[211] "211" "212" "213" "214" "215" "216" "217" "218" "219" "220"

[221] "221" "222" "223" "224" "225" "226" "227" "228" "229" "230"

[231] "231" "232" "233" "234"

*mpg’s row names are indices of the row, mtcar’s row names have car names in them*

1. If you compare the **mtcars** and **mpg** datasets, you will observe that **mtcars** does not have any manufacturer information. So let us add manufacturer information into the **mtcars** file. Use the following information for each of the 32 cars and add them into a new file, name it **mtcars1**. [Hint: Create a list with these names, and then add it as a column to the original **mtcars** file]

Mazda, Mazda, Datsun, Hornet, Hornet, Valiant, Renault, Mercedes, Mercedes, Mercedes, Mercedes, Mercedes, Mercedes, Mercedes, Cadillac, Lincoln, Chrysler, Fiat, Honda, Toyota, Toyota, Dodge, AMC, Camaro, Pontiac, Fiat, Porsche, Lotus, Ford, Ferrari, Maserati, Volvo.

#Question 3

> manufs=c("Mazda", "Mazda", "Datsun", "Hornet",

+ "Hornet", "Valiant", "Renault", "Mercedes", "Mercedes",

+ "Mercedes", "Mercedes", "Mercedes", "Mercedes", "Mercedes",

+ "Cadillac", "Lincoln", "Chrysler", "Fiat", "Honda",

+ "Toyota", "Toyota", "Dodge", "AMC", "Camaro", "Pontiac",

+ "Fiat", "Porsche", "Lotus", "Ford", "Ferrari", "Maserati",

+ "Volvo")

> mtcars1=as.data.frame(mtcars)

> mtcars1$manufacturer=manufs

* 1. How many cars are from **Mazda**? Show your results with code outputs?

#Question 3a

> library(dplyr)

> length(mtcars1$manufacturer[mtcars1$manufacturer=="Mazda"])

* 1. How many cars are from **Toyota**? Show your results with code outputs?

#Question 3b

> length(mtcars1$manufacturer[mtcars1$manufacturer=="Toyota"])

[1] 2

* 1. How many cars have displacement (**disp**) less than 200? Make a descending ordered output.

#Question 3c

> sum(mtcars1$disp<200)

[1] 16

filter(mtcars1,disp<200) %>% arrange(desc(disp))

mpg cyl disp hp drat wt qsec vs am gear carb manufacturer

1 19.2 6 167.6 123 3.92 3.440 18.30 1 0 4 4 Mercedes

2 17.8 6 167.6 123 3.92 3.440 18.90 1 0 4 4 Mercedes

3 21.0 6 160.0 110 3.90 2.620 16.46 0 1 4 4 Mazda

4 21.0 6 160.0 110 3.90 2.875 17.02 0 1 4 4 Mazda

5 24.4 4 146.7 62 3.69 3.190 20.00 1 0 4 2 Mercedes

6 19.7 6 145.0 175 3.62 2.770 15.50 0 1 5 6 Ferrari

7 22.8 4 140.8 95 3.92 3.150 22.90 1 0 4 2 Mercedes

8 21.4 4 121.0 109 4.11 2.780 18.60 1 1 4 2 Volvo

9 26.0 4 120.3 91 4.43 2.140 16.70 0 1 5 2 Porsche

10 21.5 4 120.1 97 3.70 2.465 20.01 1 0 3 1 Toyota

11 22.8 4 108.0 93 3.85 2.320 18.61 1 1 4 1 Datsun

12 30.4 4 95.1 113 3.77 1.513 16.90 1 1 5 2 Lotus

13 27.3 4 79.0 66 4.08 1.935 18.90 1 1 4 1 Fiat

14 32.4 4 78.7 66 4.08 2.200 19.47 1 1 4 1 Fiat

15 30.4 4 75.7 52 4.93 1.615 18.52 1 1 4 2 Honda

16 33.9 4 71.1 65 4.22 1.835 19.90 1 1 4 1 Toyota

* 1. Create an output file with the top 5 cars with highest displacement (**disp**) and highest engine cylinders (**cyl**). Do you notice any interesting observation within the **gear** data?

df <-mtcars1[order(-mtcars1$disp, -mtcars1$cyl),]

> head(df,5)

mpg cyl disp hp drat wt qsec vs am gear

Cadillac Fleetwood 10.4 8 472 205 2.93 5.250 17.98 0 0 3

Lincoln Continental 10.4 8 460 215 3.00 5.424 17.82 0 0 3

Chrysler Imperial 14.7 8 440 230 3.23 5.345 17.42 0 0 3

Pontiac Firebird 19.2 8 400 175 3.08 3.845 17.05 0 0 3

Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3

carb manufacturer

Cadillac Fleetwood 4 Cadillac

Lincoln Continental 4 Lincoln

Chrysler Imperial 4 Chrysler

Pontiac Firebird 2 Pontiac

Hornet Sportabout 2 Hornet

*Yes, all gears are 3.*

1. In a new contract, a company decides to export cars to Europe. The condition is that only European car manufacturers are allowed. Go back to Question (3) and check that the non-European cars are marked in RED. Now, build a new car information file with European cars only, and name it **Eurosales**.

library(sqldf)

> non\_eu=sqldf("select \* from mtcars1

+ where manufacturer in ('Mazda','Datsun','Cadillac',

+ 'Lincoln','Chrysler','Honda','Toyota','Dodge','AMC','Pontiac','Ford')")

> eu=sqldf("select \* from mtcars1

+ where manufacturer not in ('Mazda','Datsun','Cadillac',

+ 'Lincoln','Chrysler','Honda','Toyota','Dodge','AMC','Pontiac','Ford')")

> non\_eu

mpg cyl disp hp drat wt qsec vs am gear carb manufacturer

1 21.0 6 160.0 110 3.90 2.620 16.46 0 1 4 4 Mazda

2 21.0 6 160.0 110 3.90 2.875 17.02 0 1 4 4 Mazda

3 22.8 4 108.0 93 3.85 2.320 18.61 1 1 4 1 Datsun

4 10.4 8 472.0 205 2.93 5.250 17.98 0 0 3 4 Cadillac

5 10.4 8 460.0 215 3.00 5.424 17.82 0 0 3 4 Lincoln

6 14.7 8 440.0 230 3.23 5.345 17.42 0 0 3 4 Chrysler

7 30.4 4 75.7 52 4.93 1.615 18.52 1 1 4 2 Honda

8 33.9 4 71.1 65 4.22 1.835 19.90 1 1 4 1 Toyota

9 21.5 4 120.1 97 3.70 2.465 20.01 1 0 3 1 Toyota

10 15.5 8 318.0 150 2.76 3.520 16.87 0 0 3 2 Dodge

11 15.2 8 304.0 150 3.15 3.435 17.30 0 0 3 2 AMC

12 19.2 8 400.0 175 3.08 3.845 17.05 0 0 3 2 Pontiac

13 15.8 8 351.0 264 4.22 3.170 14.50 0 1 5 4 Ford

> Eurosales=eu

1. Can you add this additional information to the **mtcars1** file by a new column **Euroflag** at the end of the dataset? The value should be “Y” for European manufactured cars and rest should be “N”.

> #Question 5

> mtcars1$Euroflag[mtcars1$manufacturer %in% c('Mazda','Datsun','Cadillac',

+ 'Lincoln','Chrysler','Honda','Toyota','Dodge','AMC','Pontiac','Ford')]='N'

>

> mtcars1$Euroflag[is.na(mtcars1$Euroflag)]='Y'

> mtcars1

mpg cyl disp hp drat wt qsec vs am gear

Mazda RX4 21.0 6 160.0 110 3.90 2.620 16.46 0 1 4

Mazda RX4 Wag 21.0 6 160.0 110 3.90 2.875 17.02 0 1 4

Datsun 710 22.8 4 108.0 93 3.85 2.320 18.61 1 1 4

Hornet 4 Drive 21.4 6 258.0 110 3.08 3.215 19.44 1 0 3

Hornet Sportabout 18.7 8 360.0 175 3.15 3.440 17.02 0 0 3

Valiant 18.1 6 225.0 105 2.76 3.460 20.22 1 0 3

Duster 360 14.3 8 360.0 245 3.21 3.570 15.84 0 0 3

Merc 240D 24.4 4 146.7 62 3.69 3.190 20.00 1 0 4

Merc 230 22.8 4 140.8 95 3.92 3.150 22.90 1 0 4

Merc 280 19.2 6 167.6 123 3.92 3.440 18.30 1 0 4

Merc 280C 17.8 6 167.6 123 3.92 3.440 18.90 1 0 4

Merc 450SE 16.4 8 275.8 180 3.07 4.070 17.40 0 0 3

Merc 450SL 17.3 8 275.8 180 3.07 3.730 17.60 0 0 3

Merc 450SLC 15.2 8 275.8 180 3.07 3.780 18.00 0 0 3

Cadillac Fleetwood 10.4 8 472.0 205 2.93 5.250 17.98 0 0 3

Lincoln Continental 10.4 8 460.0 215 3.00 5.424 17.82 0 0 3

Chrysler Imperial 14.7 8 440.0 230 3.23 5.345 17.42 0 0 3

Fiat 128 32.4 4 78.7 66 4.08 2.200 19.47 1 1 4

Honda Civic 30.4 4 75.7 52 4.93 1.615 18.52 1 1 4

Toyota Corolla 33.9 4 71.1 65 4.22 1.835 19.90 1 1 4

Toyota Corona 21.5 4 120.1 97 3.70 2.465 20.01 1 0 3

Dodge Challenger 15.5 8 318.0 150 2.76 3.520 16.87 0 0 3

AMC Javelin 15.2 8 304.0 150 3.15 3.435 17.30 0 0 3

Camaro Z28 13.3 8 350.0 245 3.73 3.840 15.41 0 0 3

Pontiac Firebird 19.2 8 400.0 175 3.08 3.845 17.05 0 0 3

Fiat X1-9 27.3 4 79.0 66 4.08 1.935 18.90 1 1 4

Porsche 914-2 26.0 4 120.3 91 4.43 2.140 16.70 0 1 5

Lotus Europa 30.4 4 95.1 113 3.77 1.513 16.90 1 1 5

Ford Pantera L 15.8 8 351.0 264 4.22 3.170 14.50 0 1 5

Ferrari Dino 19.7 6 145.0 175 3.62 2.770 15.50 0 1 5

Maserati Bora 15.0 8 301.0 335 3.54 3.570 14.60 0 1 5

Volvo 142E 21.4 4 121.0 109 4.11 2.780 18.60 1 1 4

carb manufacturer Euroflag

Mazda RX4 4 Mazda N

Mazda RX4 Wag 4 Mazda N

Datsun 710 1 Datsun N

Hornet 4 Drive 1 Hornet Y

Hornet Sportabout 2 Hornet Y

Valiant 1 Valiant Y

Duster 360 4 Renault Y

Merc 240D 2 Mercedes Y

Merc 230 2 Mercedes Y

Merc 280 4 Mercedes Y

Merc 280C 4 Mercedes Y

Merc 450SE 3 Mercedes Y

Merc 450SL 3 Mercedes Y

Merc 450SLC 3 Mercedes Y

Cadillac Fleetwood 4 Cadillac N

Lincoln Continental 4 Lincoln N

Chrysler Imperial 4 Chrysler N

Fiat 128 1 Fiat Y

Honda Civic 2 Honda N

Toyota Corolla 1 Toyota N

Toyota Corona 1 Toyota N

Dodge Challenger 2 Dodge N

AMC Javelin 2 AMC N

Camaro Z28 4 Camaro Y

Pontiac Firebird 2 Pontiac N

Fiat X1-9 1 Fiat Y

Porsche 914-2 2 Porsche Y

Lotus Europa 2 Lotus Y

Ford Pantera L 4 Ford N

Ferrari Dino 6 Ferrari Y

Maserati Bora 8 Maserati Y

Volvo 142E 2 Volvo Y

1. Now set the **Euroflag** as factor.

#Question 6

> mtcars1$Euroflag=as.factor(mtcars1$Euroflag)

> str(mtcars1)

'data.frame': 32 obs. of 13 variables:

$ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...

$ cyl : num 6 6 4 6 8 6 8 4 4 6 ...

$ disp : num 160 160 108 258 360 ...

$ hp : num 110 110 93 110 175 105 245 62 95 123 ...

$ drat : num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...

$ wt : num 2.62 2.88 2.32 3.21 3.44 ...

$ qsec : num 16.5 17 18.6 19.4 17 ...

$ vs : num 0 0 1 1 0 1 0 1 1 1 ...

$ am : num 1 1 1 0 0 0 0 0 0 0 ...

$ gear : num 4 4 4 3 3 3 3 4 4 4 ...

$ carb : num 4 4 1 1 2 1 4 2 2 4 ...

$ manufacturer: chr "Mazda" "Mazda" "Datsun" "Hornet" ...

$ Euroflag : Factor w/ 2 levels "N","Y": 1 1 1 2 2 2 2 2 2 2 ...