**R PROGRAMMING**

QUES 3 : Download the Toy Dataset CSV file available on Kaggle via the link given (https://www.kaggle.com/carlolepelaars/toy-dataset). Write R commands to do the following:

a.Read the CSV into a data frame object.

CODE 🡪 mydata <- read.csv("toy\_dataset.csv")

b. Attach the data frame to the environment.

CODE 🡪 attach(mydata)

OUTPUT 🡪 The following objects are masked from mydata (pos = 3):

Age, City, Gender, Illness, Income, Number

The following objects are masked from mydata (pos = 4):

Age, City, Gender, Illness, Income, Number

c. Count the number of rows with City = New York City.

CODE 🡪 sum(City == "New York City")

OUTPUT 🡪 [1] 50307

d. Display rows that have top 5 income values.

CODE 🡪 top5 <- head(mydata[order(-Income),], 5)

top5

OUTPUT 🡪

Number City Gender Age Income Illness

109351 109351 Mountain View Male 58 177157 No

105282 105282 Mountain View Male 41 176746 No

109061 109061 Mountain View Male 61 173991 No

110878 110878 Mountain View Male 52 173826 No

112193 112193 Mountain View Male 58 172825 No

e. Find out the number of rows with income greater than average income.

CODE 🡪 avg\_income <- mean(Income)

sum(Income > avg\_income)

OUTPUT 🡪 [1] 83631

f. Find the highest salary for the female population.

CODE 🡪 max(Income[Gender == "Female"])

OUTPUT 🡪 [1] 168440

g. Randomly select 10 rows from the given dataset.

CODE 🡪 rand10 <- mydata[sample(nrow(mydata), 10),]

rand10

OUTPUT 🡪

Number City Gender Age Income Illness

32613 32613 New York City Male 52 116316 No

48980 48980 New York City Female 50 103985 No

44771 44771 New York City Female 44 95934 No

74246 74246 Los Angeles Female 48 98274 No

121829 121829 Boston Female 61 96731 No

2669 2669 Dallas Female 28 56714 Yes

55485 55485 New York City Male 42 100628 No

42191 42191 New York City Female 35 91986 No

57413 57413 New York City Female 51 76468 No

139176 139176 Austin Male 29 113331 No

h. Detach the data frame object from the environment.

CODE 🡪 detach(mydata)