NAME-SAMIKSHA
REG.NO.-12215605

ROLL NO-57

Q1)

```r
#Question 1
#LAZY LEARNER ALGORITHM(KNN ALGORITHM)

#loading necesssary libraries
library(caret)#Confusion Matrix
library(caTools)#Feature Scaling
library(class)#KNN Algorithm

data <- read.csv(file.choose())#read the CSV file
View(data)#to view data
sum(is.na(data))#to know how many na values are there in dataset
data[is.na(data)] <- 0 #to replce na values with 0
str(data)#gives structure of the data
summary(data)#gives summary of the data
data$Gender <- as.numeric(factor(data$Gender))#creating factors
data$MaritalStatus <- as.numeric(factor(data$MaritalStatus))#creating factors
normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))#formula of normalization
}
data[2:9] <- as.data.frame(lapply(data[2:9], normalize))
set.seed(123)#to randomize data
split <- sample.split(data$Product, SplitRatio = 0.7)#splitting data
train <- subset(data, split == TRUE)#training data
test <- subset(data, split == FALSE)#testing data
View(train)
View(test)
k <- 3  # Value of k used in prediction
pred <- knn(train = train[, -1], test = test[, -1], cl = train$Product, k = k)
accuracy <- mean(pred == test$Product)
cat("Accuracy: ", accuracy)
```

```
33:1  (Top Level)

Console  Terminal  Background Jobs

R 4.3.2 · ~/

 Max.   :7.000   Max.   :5.000   Max.   :104581   Max.   :360.0
> data$Gender <- as.numeric(factor(data$Gender))#creating factors
> data$MaritalStatus <- as.numeric(factor(data$MaritalStatus))#creating factors
> normalize <- function(x) {
+   return ((x - min(x)) / (max(x) - min(x)))#formula of normalization
+ }
> data[2:9] <- as.data.frame(lapply(data[2:9], normalize))
> set.seed(123)#to randomize data
> split <- sample.split(data$Product, SplitRatio = 0.7)#splitting data
> train <- subset(data, split == TRUE)#training data
> test <- subset(data, split == FALSE)#testing data
> k <- 3  # Value of k used in prediction
> pred <- knn(train = train[, -1], test = test[, -1], cl = train$Product, k = k)
> accuracy <- mean(pred == test$Product)
> cat("Accuracy: ", accuracy)
Accuracy:  0.6111111
>
```

Q2)

```r
#Question2
data <- read.csv(file.choose())
# 1
na <- sum(is.na(data))
na
# 2
na_location <- which(is.na(data),arr.ind = TRUE)
na_location
# 3
data_completecases <- data[complete.cases(data), ]
View(data_completecases)
# 4
normalize <- function(x) {(x - min(x)) / (max(x) - min(x))}
data1 <- as.data.frame(lapply(data$Product, normalize))
# 5
library(ggplot2)
ggplot(data_completecases, aes(x = Age)) + geom_histogram(binwidth = 1, fill = "blue", color = "black") +
  labs(title = "Distribution of Age", x = "Age", y = "Frequency")
```

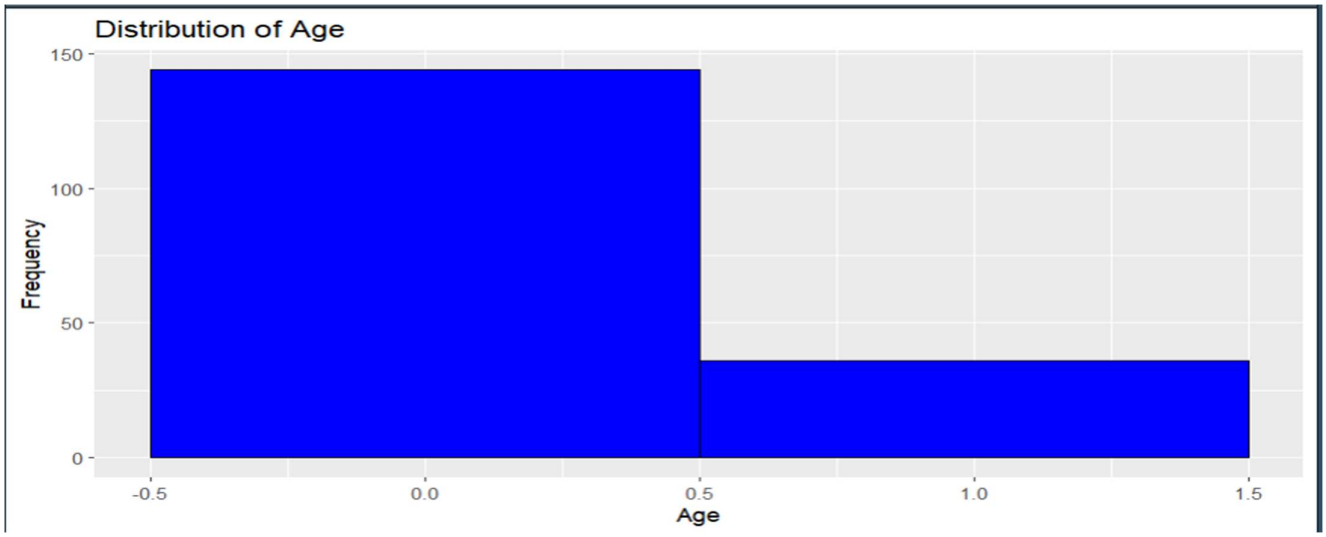
```
> # 1
> na <- sum(is.na(data))
> na
[1] 2
> # 2
> na_location <- which(is.na(data),arr.ind = TRUE)
> na_location
     row col
[1,]  19   6
[2,]  50   7
```

#3

Filter

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
1	TM195	18	Male	14	Single	3	4	29562	112
2	TM195	19	Male	15	Single	2	3	31836	75
3	TM195	19	Female	14	Partnered	4	3	30699	66
4	TM195	19	Male	12	Single	3	3	32973	85
5	TM195	20	Male	13	Partnered	4	2	35247	47
6	TM195	20	Female	14	Partnered	3	3	32973	66
7	TM195	21	Female	14	Partnered	3	3	35247	75
8	TM195	21	Male	13	Single	3	3	32973	85
9	TM195	21	Male	15	Single	5	4	35247	141
10	TM195	21	Female	15	Partnered	2	3	37521	85
11	TM195	22	Male	14	Single	3	3	36384	85
12	TM195	22	Female	14	Partnered	3	2	35247	66
13	TM195	22	Female	16	Single	4	3	36384	75
14	TM195	22	Female	14	Single	3	3	35247	75
15	TM195	23	Male	16	Partnered	3	1	38658	47
16	TM195	23	Male	16	Partnered	3	3	40932	75
17	TM195	23	Female	14	Single	2	3	34110	103
18	TM195	23	Male	16	Partnered	4	3	39795	94
20	TM195	23	Female	15	Partnered	2	2	34110	38
21	TM195	23	Male	14	Single	4	3	38658	113
22	TM195	23	Male	16	Single	4	3	40932	94
23	TM195	24	Female	16	Single	4	3	42069	94
24	TM195	24	Female	16	Partnered	5	5	44343	188
25	TM195	24	Male	14	Single	2	3	45480	113
26	TM195	24	Male	13	Partnered	3	2	42069	47
27	TM195	24	Female	16	Single	4	3	46617	75
28	TM195	25	Female	14	Partnered	3	3	48891	75
29	TM195	25	Male	14	Partnered	2	3	45480	56

#5

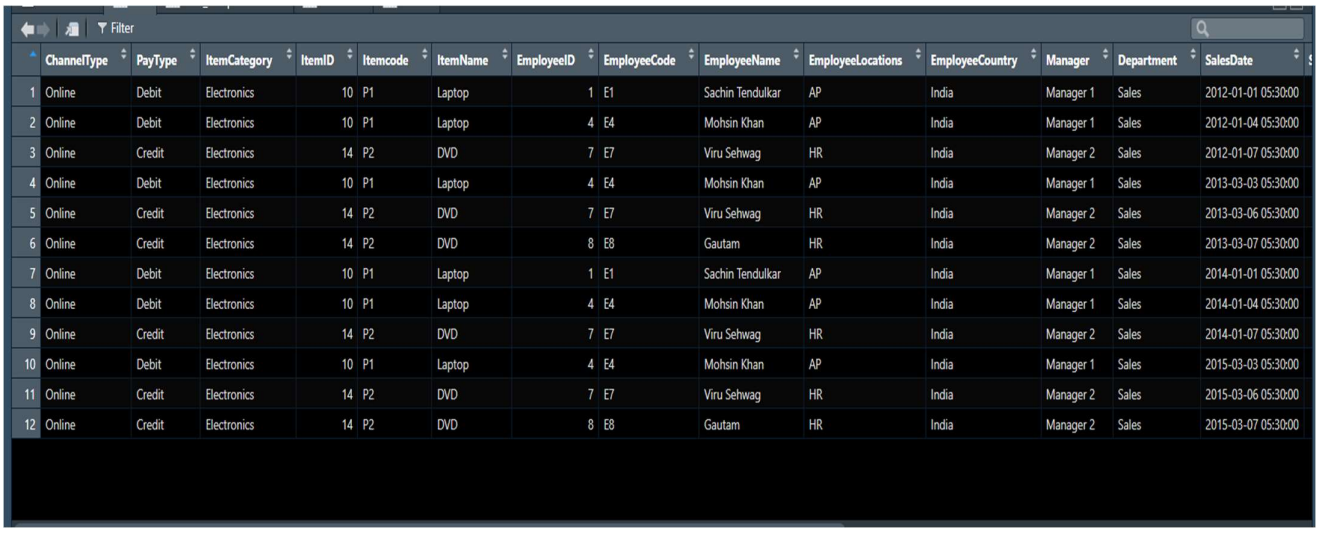


Q3)

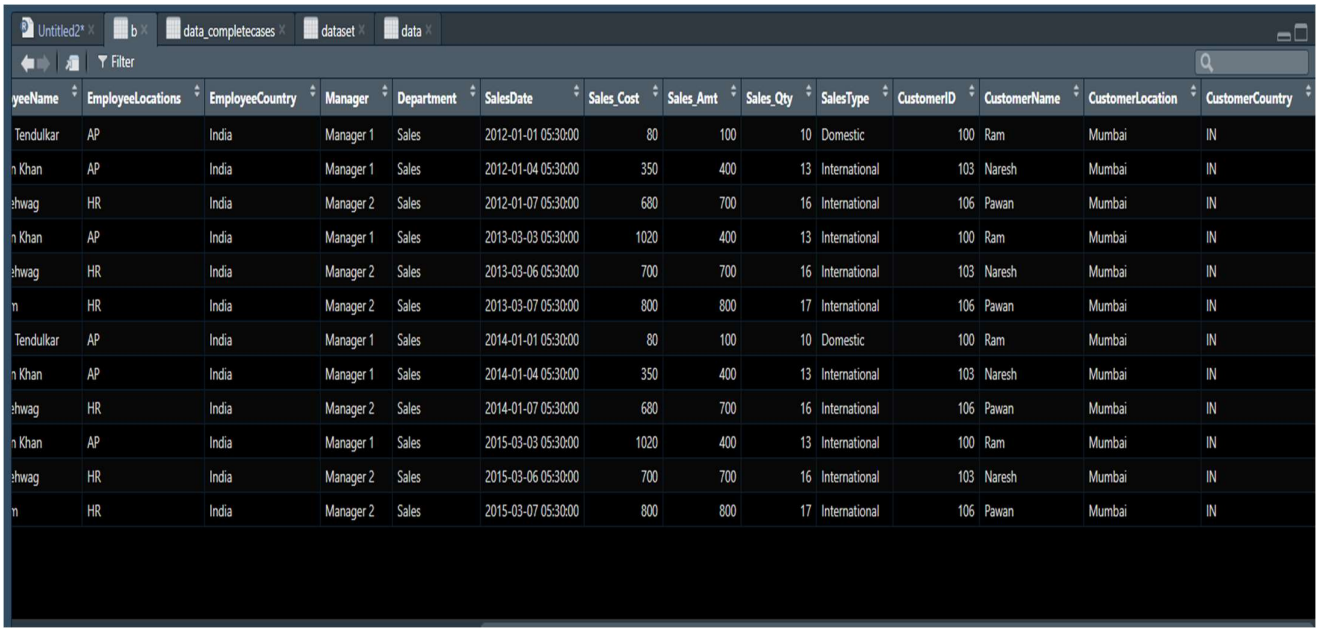
```
#Question 3
library(readxl)
library(sqldf)
dataset<-read_excel(file.choose())
View(dataset)
#1
a <- sqldf("select ItemCategory, sum(Sales_Amt) as TotalSales from dataset group by ItemCategory order by TotalSales desc LIMIT 1")
a
#2
b <- sqldf("select * from dataset where CustomerLocation = 'Mumbai' AND ChannelType = 'Online'")
View(b)
#3
c <- sqldf("select Manager, Department, sum(Sales_Amt) as TotalSales from dataset group by Manager, Department")
c
#4
d <- sqldf("select Manager, sum(Sales_Amt) as TotalSales from dataset group by Manager")
d
#5
e <- sqldf("select ItemCategory as TotalSales from dataset where ChannelType = 'Online' group by ItemCategory order by TotalSales desc LIMIT 1")
e
```

```
> #1
> a <- sqldf("select ItemCategory, sum(Sales_Amt) as TotalSales from dataset group by ItemCategory order by TotalSales desc LIMIT 1")
> a
  ItemCategory TotalSales
1     Cellular      49200
> #3
> c <- sqldf("select Manager, Department, sum(Sales_Amt) as TotalSales from dataset group by Manager, Department")
> c
    Manager Department TotalSales
1 Manager 1      Sales       4000
2 Manager 2      Sales      10400
3 Manager 3    Finance      28000
4 Manager 4 Accounting      11600
5 Manager 5 Accounting      13600
6 Manager 6      Audit       5600
7 Manager 7      Audit       6000
8 Manager 8      Audit       6000
9 Manager 9      Audit       6400
> #4
> d <- sqldf("select Manager, sum(Sales_Amt) as TotalSales from dataset group by Manager")
> d
    Manager TotalSales
1 Manager 1       4000
2 Manager 2      10400
3 Manager 3      28000
4 Manager 4      11600
5 Manager 5      13600
6 Manager 6       5600
7 Manager 7       6000
8 Manager 8       6000
9 Manager 9       6400
> #5
> e <- sqldf("select ItemCategory as TotalSales from dataset where ChannelType = 'Online' group by ItemCategory order by TotalSales desc LIMIT 1")
> e
   TotalSales
1 Electronics
>
```

2nd part



	ChannelType	PayType	ItemCategory	ItemID	Itemcode	ItemName	EmployeeID	EmployeeCode	EmployeeName	EmployeeLocations	EmployeeCountry	Manager	Department	SalesDate
1	Online	Debit	Electronics	10	P1	Laptop	1	E1	Sachin Tendulkar	AP	India	Manager 1	Sales	2012-01-01 05:30:00
2	Online	Debit	Electronics	10	P1	Laptop	4	E4	Mohsin Khan	AP	India	Manager 1	Sales	2012-01-04 05:30:00
3	Online	Credit	Electronics	14	P2	DVD	7	E7	Viru Sehwag	HR	India	Manager 2	Sales	2012-01-07 05:30:00
4	Online	Debit	Electronics	10	P1	Laptop	4	E4	Mohsin Khan	AP	India	Manager 1	Sales	2013-03-03 05:30:00
5	Online	Credit	Electronics	14	P2	DVD	7	E7	Viru Sehwag	HR	India	Manager 2	Sales	2013-03-06 05:30:00
6	Online	Credit	Electronics	14	P2	DVD	8	E8	Gautam	HR	India	Manager 2	Sales	2013-03-07 05:30:00
7	Online	Debit	Electronics	10	P1	Laptop	1	E1	Sachin Tendulkar	AP	India	Manager 1	Sales	2014-01-01 05:30:00
8	Online	Debit	Electronics	10	P1	Laptop	4	E4	Mohsin Khan	AP	India	Manager 1	Sales	2014-01-04 05:30:00
9	Online	Credit	Electronics	14	P2	DVD	7	E7	Viru Sehwag	HR	India	Manager 2	Sales	2014-01-07 05:30:00
10	Online	Debit	Electronics	10	P1	Laptop	4	E4	Mohsin Khan	AP	India	Manager 1	Sales	2015-03-03 05:30:00
11	Online	Credit	Electronics	14	P2	DVD	7	E7	Viru Sehwag	HR	India	Manager 2	Sales	2015-03-06 05:30:00
12	Online	Credit	Electronics	14	P2	DVD	8	E8	Gautam	HR	India	Manager 2	Sales	2015-03-07 05:30:00



yeeName	EmployeeLocations	EmployeeCountry	Manager	Department	SalesDate	Sales_Cost	Sales_Amt	Sales_Qty	SalesType	CustomerID	CustomerName	CustomerLocation	CustomerCountry
Tendulkar	AP	India	Manager 1	Sales	2012-01-01 05:30:00	80	100	10	Domestic	100	Ram	Mumbai	IN
n Khan	AP	India	Manager 1	Sales	2012-01-04 05:30:00	350	400	13	International	103	Naresh	Mumbai	IN
ehwag	HR	India	Manager 2	Sales	2012-01-07 05:30:00	680	700	16	International	106	Pawan	Mumbai	IN
n Khan	AP	India	Manager 1	Sales	2013-03-03 05:30:00	1020	400	13	International	100	Ram	Mumbai	IN
ehwag	HR	India	Manager 2	Sales	2013-03-06 05:30:00	700	700	16	International	103	Naresh	Mumbai	IN
n	HR	India	Manager 2	Sales	2013-03-07 05:30:00	800	800	17	International	106	Pawan	Mumbai	IN
Tendulkar	AP	India	Manager 1	Sales	2014-01-01 05:30:00	80	100	10	Domestic	100	Ram	Mumbai	IN
n Khan	AP	India	Manager 1	Sales	2014-01-04 05:30:00	350	400	13	International	103	Naresh	Mumbai	IN
ehwag	HR	India	Manager 2	Sales	2014-01-07 05:30:00	680	700	16	International	106	Pawan	Mumbai	IN
n Khan	AP	India	Manager 1	Sales	2015-03-03 05:30:00	1020	400	13	International	100	Ram	Mumbai	IN
ehwag	HR	India	Manager 2	Sales	2015-03-06 05:30:00	700	700	16	International	103	Naresh	Mumbai	IN
n	HR	India	Manager 2	Sales	2015-03-07 05:30:00	800	800	17	International	106	Pawan	Mumbai	IN