1.Randomly Sample the iris dataset such as 80% data for training and 20% for test and   create

Logistics regression with train data, use species as target and petals width and

length as feature variables , Predict the probability of the model using test data,  Create

Confusion matrix for above test model

2. (i)Write suitable R code to compute the mean, median ,mode of the following values

            c(90, 50, 70, 80, 70, 60, 20, 30, 80, 90, 20)

  (ii) Write R code to find 2nd  highest and 3 rd Lowest value of above problem.

3. Explore the airquality dataset. It contains daily air quality measurements from New York during

a period of five months:

• Ozone: mean ozone concentration (ppb), • Solar.R: solar radiation (Langley),

• Wind: average wind speed (mph), • Temp: maximum daily temperature in degrees Fahrenheit,

• Month: numeric month (May=5, June=6, and so on),• Day: numeric day of the month (1-

4).

i. Compute the mean temperature(don’t use build in function)

ii.Extract the first five rows from airquality.

iii.Extract all columns from airquality except Temp and Wind

iv.Which was the coldest day during the period?

v.How many days was the wind speed greater than 17 mph?

5. (i)Get the Summary Statistics of air quality dataset

(ii)Melt airquality data set and display as a long – format data?

(iii)Melt airquality data and specify month and day to be “ID variables”?

 (iv)Cast the molten airquality data set with respect to month and date features

 (v) Use cast function appropriately and compute the average of Ozone, Solar.R , Wind and

              temperature per month?

6.(i) Find any missing values(na) in features and drop the missing values if its less than 10%

else replace that with  mean of that feature.

   (ii) Apply a linear regression algorithm using Least Squares Method on “Ozone” and “Solar.R”

   (iii)Plot Scatter plot between Ozone and Solar and add regression line created by above

model

7. Load dataset named ChickWeight,

( i).Order the data frame, in ascending order by feature name “weight” grouped by   feature

“diet” and Extract the last 6 records from order data frame.

 (ii).a Perform melting function based on “Chick&quot;, &quot;Time&quot;, &quot;Diet&quot;   features as ID variables

 b. Perform cast function to display the mean value of weight grouped by Diet

 c. Perform cast function to display the mode of weight grouped by Diet

8. a.  Create Box plot for “weight” grouped by “Diet”

          b. Create a Histogram for “weight” features belong to Diet- 1 category

          c.  Create Scatter plot for “ weight” vs “Time” grouped by Diet

9.   a. Create multi regression model to find a weight of the chicken , by “Time” and “Diet” as  as

predictor variables

          b. Predict weight for Time=10 and Diet=1

           c. Find the error in model for same

10 .For this exercise, use the (built-in) dataset Titanic.

   a. Draw a Bar chart to show details of “Survived” on the Titanic based on passenger Class

   b. Modify the above plot based on gender of people who survived

  c. Draw histogram plot to show distribution of feature “Age”

11. Explore the USArrests dataset, contains the number of arrests for murder, assault, and rape

for each of the 50 states in 1973. It also contains the percentage of people in the state who live in

an urban area.

(i) a. Explore the summary of Data set, like number of Features and its type. Find the number

        of records for each feature. Print the statistical feature of data

         b. Print the state which saw the largest total number of rape

         c. Print the states with the max &amp; min crime rates for murder

   (ii).a. Find the correlation among the features

    b. Print the states which have assault arrests more than median of the country

    c. Print the states are in the bottom 25% of murder

  (iii). a. Create a histogram and density plot of murder arrests by US stat

b. Create the plot that shows the relationship between murder arrest rate and  proportion

of the population that is urbanised by state. Then enrich the chart by adding assault

arrest rates (by colouring the points from blue (low) to red (high)).

        c. Draw a bar graph to show the murder rate for each of the 50 states .

12.. a. Create a data frame based on below table.

Month 1 2 3 4 5 6 7 8 9 10 11 12

Spends 1000 4000 5000 4500 3000 4000 9000 11000 15000 12000 7000 3000

Sales 9914 40487 54324 50044 34719 42551 94871 118914 158484 131348 78504 36284

b. Create a regression model for that data frame table to show the amount of sales(Sales) based on

the how much the company spends (Spends) in advertising

c. Predict the Sales if Spend=13500

1.      i) Write a function called kelvin\_to\_celsius() that takes a  
 temperature in Kelvin and returns   that  
 temperature in Celsius (**Hint:**To convert from Kelvin to Celsius you  
 subtract 273.15)  
  
  
 (ii) Write suitable R code to compute the  
 mean, median ,mode of the following values  
  
            c(90, 50, 70, 80, 70, 60, 20, 30,  
 80, 90, 20)        
  
  
     (iii) Write R code to  
 find 2nd  highest and 3rd  
 Lowest value of above problem.

A) #1st

kelvin\_to\_celsius <- function(temp\_c) {

temp\_k <- (temp\_c - 273.15)

return(temp\_c)

}

kelvin\_to\_celsius(0)

#2nd

v <- c(90, 50, 70, 80, 70, 60, 20, 30,

80, 90, 20)

print(v)

mean(v)

median(v)

getmode <- function(v)

{

uniqv <-unique(v)

uniqv[which.max(tabulate(match(v,uniqv)))]

}

result <- getmode(v)

print(result)

#3rd

x <- c(90, 50, 70, 80, 70, 60, 20, 30,

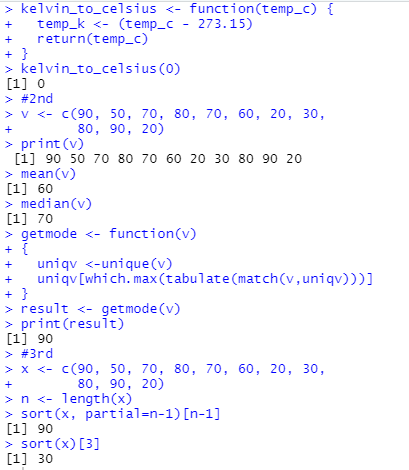
80, 90, 20)

n <- length(x)

sort(x, partial=n-1)[n-1]

sort(x)[3]

OUTPUT



2.      Explore the airquality dataset. It  
contains daily air quality measurements from New York during a period of five  
months:  
  
• Ozone: mean  
ozone concentration (ppb),  
  
• Solar.R:  
solar radiation (Langley),  
  
• Wind:  
average wind speed (mph),  
  
• Temp:  
maximum daily temperature in degrees Fahrenheit,  
  
• Month:  
numeric month (May=5, June=6, and so on),  
  
• Day:  
numeric day of the month (1-31).                                                              
  
  
  
i. Compute the mean  
temperature(don’t use build in function)  
  
ii.  
Extract the first five rows from airquality.  
  
  
iii.  
Extract all columns from airquality except  
Temp and Wind  
  
iv.  
Which was the coldest day during the period?  
  
v.  
How many days was the wind speed greater than 17 mph?

A)airqdata<-airquality

print(airqdata)

#question1

mean = sum(airqdata$Temp)/length(airqdata$Temp)

print(mean)

#question2

airqdata[(1:5),]

#question3

airqdata[,c("Temp")]=NULL

airqdata[,c("Wind")]=NULL

print(airqdata)

#question4

airqdata<-airquality

print(airqdata)

cold<-min(airqdata$Temp)

print(cold)

print(match(56,airqdata$Temp))

airqdata[(match(56,airqdata$Temp)),6]

#question5

airqdata<-airquality

print(airqdata)

w<-(airqdata$Wind)>17

print(w)

print(nrow(w))

OUTPUT

