Northeastern University

Toronto, Canada

A Report on

Executive Summary of Module 1

Introduction to Data Analytics (ALY6000)

Guided by:

Prof. Mohammad Shafiqul Islam

Submitted by:

Name : Pratikkumar Indravadan Malaviya

NUID : 002963548

Date of submission : 20th January'2022

INDEX

a) A scatter plot of the Sales ~ temp data

```
#4 Plot a sales ~ temp scatter plot using the data below
Sales = c(7,11,15,20,19,11,18,10,6,22)
Temperature = c(69,81,77,84,80,97,87,70,65,90)
plot(x=Temperature,y=Sales)
```
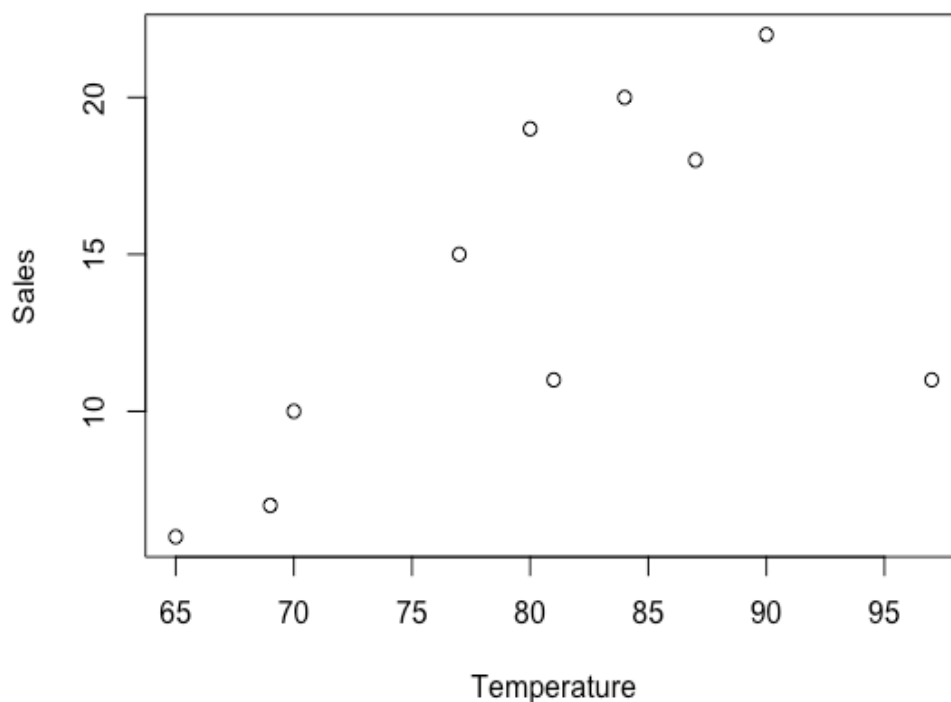

Fig 1. Sales ~ Temperature

b) The Mean Temperature

```
#5 Find the mean temperature
Temperature = c(69,81,77,84,80,97,87,70,65,90)
a = mean(Temperature)
print(a)

## [1] 80
```

c) Display the steps after 6 and 7

```
#6 Delete the 3rd element from the sales vector
Sales = c(7,11,15,20,19,11,18,10,6,22)
b = Sales[c(-3)]
print(b)

## [1]  7 11 20 19 11 18 10  6 22

#7 Insert 16 as the 3rd element into the sales vector
inserting_element = append(Sales,16,2)
print(inserting_element)

##  [1]  7 11 16 15 20 19 11 18 10  6 22
```

d) Display the names vector

```
#8 Create a vector <names> with elements Tom, Dick, Harry
names = c('Tom','Dick','Harry')
print(names)

## [1] "Tom"   "Dick"  "Harry"
```

e) Display the 5 row by 2 column of 10 integers

```
#9 Create a 5 row and 2 column matrix of 10 integers
SampleMatrix = matrix(c('3','6','7','8','4','2','8','1','9','7'),
                  nrow = 5,
                  ncol = 2,
                  byrow = TRUE)
print(SampleMatrix)

##      [,1] [,2]
## [1,] "3"  "6"
## [2,] "7"  "8"
## [3,] "4"  "2"
## [4,] "8"  "1"
## [5,] "9"  "7"
```

f) Display the icSales data frame

```
#10 Create a data frame <icSales> with sales and temp attributes
Sales = c(7,11,15,20,19,11,18,10,6,22)
Temperature = c(69,81,77,84,80,97,87,70,65,90)
icSales  = data.frame(Sales,Temperature)
#11 Displaying structure
print(icSales)

##   Sales Temperature
## 1     7          69
## 2    11          81
## 3    15          77
```

```
## 4      20           84
## 5      19           80
## 6      11           97
## 7      18           87
## 8      10           70
## 9       6           65
## 10     22           90
```

g) Display the summery of the icSales data frame

```
Sales = c(7,11,15,20,19,11,18,10,6,22)
Temperature = c(69,81,77,84,80,97,87,70,65,90)
icSales  = data.frame(Sales,Temperature)
print(summary(icSales))

##      Sales           Temperature
##  Min.   : 6.00   Min.   :65.00
##  1st Qu.:10.25   1st Qu.:71.75
##  Median :13.00   Median :80.50
##  Mean   :13.90   Mean   :80.00
##  3rd Qu.:18.75   3rd Qu.:86.25
##  Max.   :22.00   Max.   :97.00
```

h) Display the variables only from the Student.csv data set

```
#14 Displaying variable names
print(names(FileData))

## [1] "StudentID"      "First"          "Last"           "Math"
## [5] "Science"        "Social.Studies"
```

i) A summary of the information you learned about the data set based on the instruction you followed.

- The final scatter plot is based on two given vector values, Sales and Temperature. Moreover, the plot itself represents a mathematical fluctuation that has been generated by the two variables for a set of data, where the position of the x-axis and y-axis has been taken by Temperature and Sales respectively.

- Furthermore, a sample mean is calculated of the Temperature vector by applying "mean()" function (It is an arithmetic average of data set and calculated by taking the sum of all data points, then divided by the total number of integers) and it resulted in 81.

- A matrix is created by the Matrices-data type of R language. The mode of the matrix is 5 rows by 2 columns with 10 integer values.

- Another data type is known as a data frame which represents data in tabular form. Note that, data frames could contain multiple columns with different data types (if the first column is numeric then the second column can be character and the third can be logical), where we could actually name the column with an appropriate name and enter values row-wise. The data frames are created by "data.frame()" function.

- The summery() function can be used to a summarized data set. In the given example of sales and temperature, it summarises
  - minimum value as "Min",
  - value of the first quartile as "1st Qu",
  - the median value as "Median",
  - value of the third quartile as "3rd Qu",
  - maximum value as "Max".

- Lastly, the .csv file is an external data source that can be imported into R by read.csv("file_path") function. In the function argument, we need to define the actual path of the file where it is stored, then the R can able to read all data from the files and becomes ready to perform appropriate tasks on the data.

Bibliography

1. R Vectors - Create, Access, Modify and Delete a Vector Elements in R,
   http://www.awasthiashish.com/2018/08/r-vectors-create-access-modify-and-delete-a-vector-elements-in-r.html

2. Tutorialspoint.com. 2022. *R Tutorial*. *https://www.tutorialspoint.com/r/index.htm*

3. Append in R: How to Add Elements to a Vector in R, *https://r-lang.com/append-in-r/*

4. Kabacoff, Robert. R In Action: Data Analysis and Graphics with R. Manning, 2015.

Appendix

```r
#1 Display Name
print("Pratikkumar Malaviya")

#2 Install the vcd package
install.packages("vcd")

#3 Import the vcd library
library(vcd)

#4 Plot a sales ~ temp scatter plot using the data below
Sales = c(7,11,15,20,19,11,18,10,6,22)
Temperature = c(69,81,77,84,80,97,87,70,65,90)
plot(x= Temperature,y=Sales)

#5 Find the mean temperature
Temperature = c(69,81,77,84,80,97,87,70,65,90)
a = mean(Temperature)
print(a)

#6 Delete the 3rd element from the sales vector
Sales = c(7,11,15,20,19,11,18,10,6,22)
b = Sales[c(-3)]
print(b)

#7 Insert 16 as the 3rd element into the sales vector
inserting_element = append(Sales,16,2)
print(inserting_element)

#8 Create a vector <names> with elements Tom, Dick, Harry
names = c('Tom','Dick','Harry')
print(names)

#9 Create a 5 row and 2 column matrix of 10 integers
SampleMatrix = matrix(c('3','6','7','8','4','2','8','1','9','7'),
          nrow = 5,
          ncol = 2,
          byrow = TRUE)
print(SampleMatrix)

#10 Create a data frame <icSales> with sales and temp attributes
Sales = c(7,11,15,20,19,11,18,10,6,22)
Temperature = c(69,81,77,84,80,97,87,70,65,90)
icSales  = data.frame(Sales,Temperature)

#11 Displaying structure
print(icSales)

#12 Displaying Summery <icscale>
```

```
print(summary(icSales))

#13 importing csv file
FileData = read.csv("/Users/pratik_4511/Downloads/Student.csv",
          header = TRUE,
          stringsAsFactors = FALSE)

#14 Displaying variable names
print(names(FileData))
```

Github

Username : pratik4511
Repository : https://github.com/pratik4511/malaviya_m1_project-1