

**Date: 17 Jan 2022**

## **Soni\_M1\_Project1**

Submitted by

Soni Manan

Student Id: 002982645

Under the guidance of

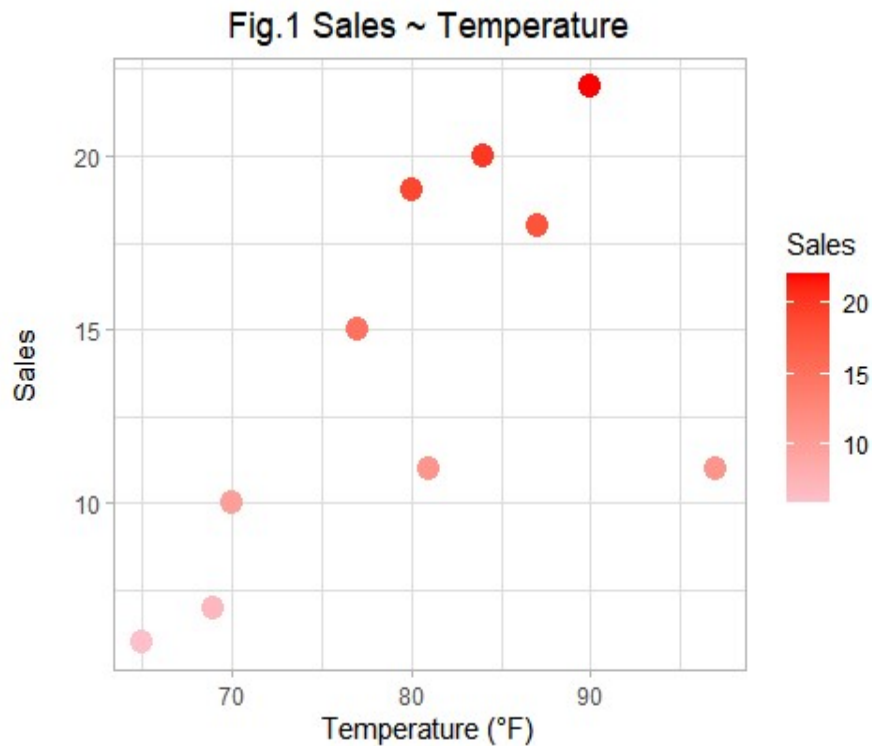
**Prof. Dr. Shafiqul**

**Northeastern University, Toronto**

Master of Professional Studies in Analytics, College of Professional Studies.

### a. A scatter plot of the Sales ~ Temperature data

Scatter plot based on provided data of sales and temperature is shown(Fig.1) below which clearly illustrates that sales during range of temperature comparatively good.



### b. The mean of temperature

Mean of the temperature is 80.

### c. The data after updating sales vector

There was a change in Data that actual sales were 16 instead 15. After updating sales, decision becomes clearer.

### d. Displaying the <names> vector

```
names <- c("Tom", "Dick", "Harry")  
## "Tom" "Dick" "Harry"
```

### e. The 5x2 Matrix of 10 Integers

Integers starting from 1 to 10 has aligned by 5 rows.

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

### f. The icSales data frame

For, better Observation, Sales and Temperature have combined into a data frame name icSales which is shown below.

	Sales	Temperature
1	7	69
2	11	81
3	16	77
4	20	84
5	19	80
6	11	97
7	18	87
8	10	70
9	6	65
10	22	90

### g. The summary of the icSales data frame.

This summary denotes Means, Medians, Minimums and Maximums of Sales and Temperature.

Sales		Temperature	
Min.	: 6.00	Min.	:65.00
1st Qu.:	10.25	1st Qu.:	71.75
Median	:13.50	Median	:80.50
Mean	:14.00	Mean	:80.00
3rd Qu.:	18.75	3rd Qu.:	86.25
Max.	:22.00	Max.	:97.00

### h. The variables from the Student.csv.

*StudentID, First (Name), Last (Name), Math, Science and Social Studies* are variables of **Student.csv** dataset.

## **i. Summary**

After observing First Dataset of Sales & Temperature and plotting its vectors on scatter graph it seems like temperature plays important role in sales. Whenever temperature falls below 70 F or rise above 95 F sales drops. On the contrary, temperature between 70 to 90 F is much suitable for sales and slowing down production/manufacturing in such temperature can be more beneficial.

Focusing on second data set of students, it seems like one of student named Jane Weary is absent during her science exam. However, she performs well in other two. In addition, other three perform very well. Student named Mary O'Leary seems like has topped among them. Dan Thornton III is an average student who has scored B or B- in each subject.

## Bibliography

Tutorialspoint. R tutorial, <https://www.tutorialspoint.com/r/index.htm>

IRTFM. "How to Insert an Element in the Middle of a List?" *Stack Overflow*, Jul 26 '21 at 17:04, <https://stackoverflow.com/questions/25025695/how-to-insert-an-element-in-the-middle-of-a-list>.

Abhijit Balki. "Get Column Name of CSV file" *Stack Overflow*, May 20'19 at 10:05, <https://stackoverflow.com/questions/56217549/get-column-name-of-csv-file>

Wickham, Hadley, et al. "Create Elegant Data Visualizations Using the Grammar of Graphics." *Create Elegant Data Visualizations Using the Grammar of Graphics* • ggplot2, RStudio, <https://ggplot2.tidyverse.org/index.html>.

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

## Appendix

```
# check if package exist and installing package 'vcd'
if(!require('vcd',character.only = TRUE)) install.packages('vcd')
# importing 'vcd'
library(vcd)

# creating Vectors
sales <- c(7,11,15,20,19,11,18,10,6,22)
temperature = c(69,81,77,84,80,97,87,70,65,90)

# plotting scatter graph for sales ~ temperature
library(ggplot2)
ggplot(data.frame(sales,temperature), aes(x = temperature , y=sales , color =
..y..)) +
  geom_point(size=4) +
  labs(
    title = "Fig.1 Sales ~ Temperature", # adds title
    x = "Temperature (°F)", # x-axis label
    y = "Sales\n", # y-axis label
    color = "Sales" # color legend
  ) + scale_color_gradient(low="pink", high="red") + theme_light() +
  theme(plot.title = element_text(hjust = 0.5))

meanOfTemperature <- mean(temperature) # Mean of temperature Vector

# Removing From Index which is 3 as Indexing starts from 1 in 'R'
sales <- c(sales[-3])
# Insert 16 at index 2
sales <- append(sales,16,2)

# allocationg Vector of strings in <names> Vector
names <- c("Tom","Dick","Harry")

mMatrix <- matrix(c(1:10),nrow=5,byrow=FALSE) # creating 5x2 Matrix

# creating dataFrame from sales, temperature named <icSales>
icSales <- data.frame('Sales'= sales,'Temperature' = temperature)
print(str(icSales))

# finding summary of Data Frame <icSales>
mSummary = summary(icSales)

# to open Student.csv file read.csv is built in function provided by 'R'
student_data = read.csv("./Student.csv") # change CWD for this command

# two functions names() and col_names() which are built-in in 'R'
names(student_data) # Printing Variables of Student.csv
```