

SHETH L.U.J & SIR M.V COLLEGE

(DATA ANALYSIS WITH SAS/SPSS/R)

OUTPUTS-

PRAC 1-

RStudio interface showing R code and environment pane.

```
R - R 4.5.2 - C:/tanvis087/
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
R - R 4.5.2 - C:/tanvis087/
> df1 <- read.csv("diabetes_012_health_indicators_.csv")
> num1 <- df1[, sapply(df1, is.numeric)][, 1]
> summary(num1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000  0.0000  0.0000  0.2969  0.0000  2.0000
>
> library(psych)
> describe(num1)
      vars      n mean  sd median trimmed mad min max range skew kurtosis se
X1      1 253680  0.3  0.7    0   0.12   0   0   2    2  1.98   1.98  0
> |
```

Environment pane:

Object	Class	Size
df	data.frame	7 obs. of 4 variables
df1	data.frame	253680 obs. of 22 variables
df2	data.frame	7 obs. of 4 variables
num1	numeric	Large numeric (253680 elements, 2 MB)

Files pane:

Name	Size	Modified
..		
.Rhstory	12.6 KB	Dec 15, 2025, 11:07 AM
online_shopping.csv	8.7 MB	Dec 15, 2025, 11:00 AM

PRAC 2-

RStudio interface showing R code and environment pane.

```
R - R 4.5.2 - C:/tanvis087/
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
R - R 4.5.2 - C:/tanvis087/
> df2 <- read.csv("Best Selling Mobile Phones 2020.csv")
> cat_col <- names(df2)[sapply(df2, is.character)][1]
> table(df2[[cat_col]])
      Galaxy A11      Galaxy A21s
Galaxy s20, s20+, s20 Ultra  1
      iPhone 12 and iPhone 12 Mini  1
      iPhone 12 Pro Max      iPhone SE 2nd generation  1
      Redmi Note 9 Pro      1
      1
>
> library(dplyr)
> df2 %>% count(.data[[cat_col]])
  Phone n
1 Galaxy A11 1
2 Galaxy A21s 1
3 Galaxy s20, s20+, s20 Ultra 1
4 Redmi Note 9 Pro 1
5 iPhone 12 Pro Max 1
6 iPhone 12 and iPhone 12 Mini 1
7 iPhone SE 2nd generation 1
> |
```

Environment pane:

Object	Class	Size
df	data.frame	7 obs. of 4 variables
df1	data.frame	253680 obs. of 22 variables
df2	data.frame	7 obs. of 4 variables
cat_col	character	"Phone"
cat2	character	chr [1:7] " iPhone 12 and iPhone 12 Mini " " Galax..
num1	numeric	Large numeric (253680 elements, 2 MB)

Files pane:

Name	Size	Modified
..		
.Rhstory	12.6 KB	Dec 15, 2025, 11:07 AM
online_shopping.csv	8.7 MB	Dec 15, 2025, 11:00 AM

SHETH L.U.J & SIR M.V COLLEGE

(DATA ANALYSIS WITH SAS/SPSS/R)

PRAC 3-

RStudio interface showing R code execution. The code reads a CSV file named 'class.csv' and displays the resulting data frame 'df3'. The data is categorized by animal class (Amphibian, Bird, Bug, Fish, Invertebrate, Mammal, Reptile) and animal names (aardvark, antelope, bear, boar, buffalo, calf, cavy, cheetah, deer, dolphin, elephant, fruitbat, giraffe, girl, goat, gorilla, hamster, hare, leopard, lion, lynx, mink, mole, mongoose, opossum, oryx, platypus, polecat, pony, porpoise, puma, pussy cat, raccoon, reindeer, seal, sealion, squirrel, vampire, vole, wallaby, wolf).

```
R - R 4.5.2 - C:/tanvis087/
> df3 <- read.csv("class.csv")
> cat_cols <- names(df3)[sapply(df3, is.character)]
> table(df3[[cat_cols[1]]], df3[[cat_cols[2]])
```

Environment pane shows the Global Environment with variables: df (7 obs. of 4 variables), df1 (253680 obs. of 22 variables), df2 (7 obs. of 4 variables), and df3 (7 obs. of 4 variables). The Files pane shows the current directory (C:/tanvis087) containing files: .Rhistory (12.6 KB, Dec 15, 2025, 11:07 AM) and online_shopping.csv (8.7 MB, Dec 15, 2025, 11:00 AM).

RStudio interface showing R code execution. The code reads a CSV file named 'class.csv' and displays the resulting data frame 'df3'. The data is categorized by animal class (Amphibian, Bird, Bug, Fish, Invertebrate, Mammal, Reptile) and animal names (chicken, crow, dove, duck, flamingo, gull, hawk, kiwi, lark, ostrich, parakeet, penguin, pheasant, rhea, skimmer, skua, sparrow, swan, vulture, wren).

```
R - R 4.5.2 - C:/tanvis087/
Bug
0
Fish
1
Invertebrate
0
Mammal
0
Reptile
0
```

Environment pane shows the Global Environment with variables: df (7 obs. of 4 variables), df1 (253680 obs. of 22 variables), df2 (7 obs. of 4 variables), and df3 (7 obs. of 4 variables). The Files pane shows the current directory (C:/tanvis087) containing files: .Rhistory (12.6 KB, Dec 15, 2025, 11:07 AM) and online_shopping.csv (8.7 MB, Dec 15, 2025, 11:00 AM).

SHETH L.U.J & SIR M.V COLLEGE

(DATA ANALYSIS WITH SAS/SPSS/R)

The RStudio interface displays a script in the Source pane with the following code:

```
Fish
0
Invertebrate
1
Mammal
0
Reptile
0

flea, gnat, honeybee, housefly, ladybird, moth, termite, wasp
Amphibian 1
Bird 0
Bug 1
Fish 0
Invertebrate 0
Mammal 0
Reptile 0

frog, frog, newt, toad
Amphibian 1
Bird 0
Bug 0
Fish 0
Invertebrate 0
Mammal 0
Reptile 0

pitviper, seasnake, slowworm, tortoise, tuatara
Amphibian 0
Bird 0
Bug 0
Fish 0
Invertebrate 0
Mammal 0
Reptile 1
```

The Environment pane on the right shows the following data objects:

Object	Obs.	Vars.
df	7	4
df1	253680	22
df2	7	4
df3	7	4

The Values pane shows the following data:

cat_col	cat_cols
"Phone"	chr [1:2] "Class_Type" "Animal_Names"

PRAC-4

The RStudio interface displays a script in the Source pane with the following code:

```
> df4 <- read.csv("mental_health_digital_behavior_data.csv")
>
> num_col <- names(df4)[sapply(df4, is.numeric)][1]
>
> t.test(df4[[num_col]], mu = mean(df4[[num_col]]))

One Sample t-test

data: df4[[num_col]]
t = 0, df = 499, p-value = 1
alternative hypothesis: true mean is not equal to 360.4378
95 percent confidence interval:
 355.2729 365.6027
sample estimates:
mean of x
 360.4378
```

The Environment pane on the right shows the following data objects:

Object	Obs.	Vars.
df2	7	4
df3	7	4
df4	500	9

The Values pane shows the following data:

cat_col	cat_cols	cat2	num_col
"Phone"	chr [1:2] "Class_Type" "Animal_Names"	chr [1:7] " iPhone 12 and iPhone 12 Mini "	"daily_screen_time_min"

SHETH L.U.J & SIR M.V COLLEGE

(DATA ANALYSIS WITH SAS/SPSS/R)

PRAC 5-

RStudio interface showing a Welch Two Sample t-test. The source editor contains the following R code:

```
R - R 4.5.2 - C:/tanvis087/
> df5 <- read.csv("online_shopping.csv")
> num_col <- names(df5)[sapply(df5, is.numeric)][1]
> cat_col <- names(df5)[sapply(df5, is.character)][1]
> df_temp <- df5[, c(num_col, cat_col)]
> df_temp[[cat_col]] <- as.factor(df_temp[[cat_col]])
> df_temp <- df_temp[df_temp[[cat_col]] %in% levels(df_temp[[cat_col]])[1:2], ]
> t.test(df_temp[[num_col]] ~ df_temp[[cat_col]])
```

The console output shows the results of the Welch Two Sample t-test:

```
Welch Two Sample t-test

data: df_temp[[num_col]] by df_temp[[cat_col]]
t = 310.65, df = 33026, p-value < 2.2e-16
alternative hypothesis: true difference in means between group and group F is not equal to 0
95 percent confidence interval:
 25797.97 26125.58
sample estimates:
mean in group mean in group F
 52939.00      26977.22
```

The Environment pane shows the following data frames:

Object	Variables	Observations
df3	4	7
df4	9	500
df5	21	52955

The Files pane shows the following files:

Name	Size	Modified
.Rhistory	12.6 KB	Dec 15, 2025, 11:07 AM
online_shopping.csv	8.7 MB	Dec 15, 2025, 11:00 AM

PRAC 6-

RStudio interface showing a Paired t-test. The source editor contains the following R code:

```
R - R 4.5.2 - C:/tanvis087/
> df6 <- read.csv("emails.csv")
> nums6 <- df6[, sapply(df6, is.numeric)]
> t.test(nums6[,1], nums6[,2], paired = TRUE)
```

The console output shows the results of the Paired t-test:

```
Paired t-test

data: nums6[, 1] and nums6[, 2]
t = 5.2872, df = 5171, p-value = 1.293e-07
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 0.2846774 0.6201950
sample estimates:
mean difference
 0.4524362
```

The Environment pane shows the following data frames:

Object	Variables	Observations
df5	21	52955
df6	3002	5172
nums6	3001	5172

The Files pane shows the following files:

Name	Size	Modified
.Rhistory	12.6 KB	Dec 15, 2025, 11:07 AM
online_shopping.csv	8.7 MB	Dec 15, 2025, 11:00 AM