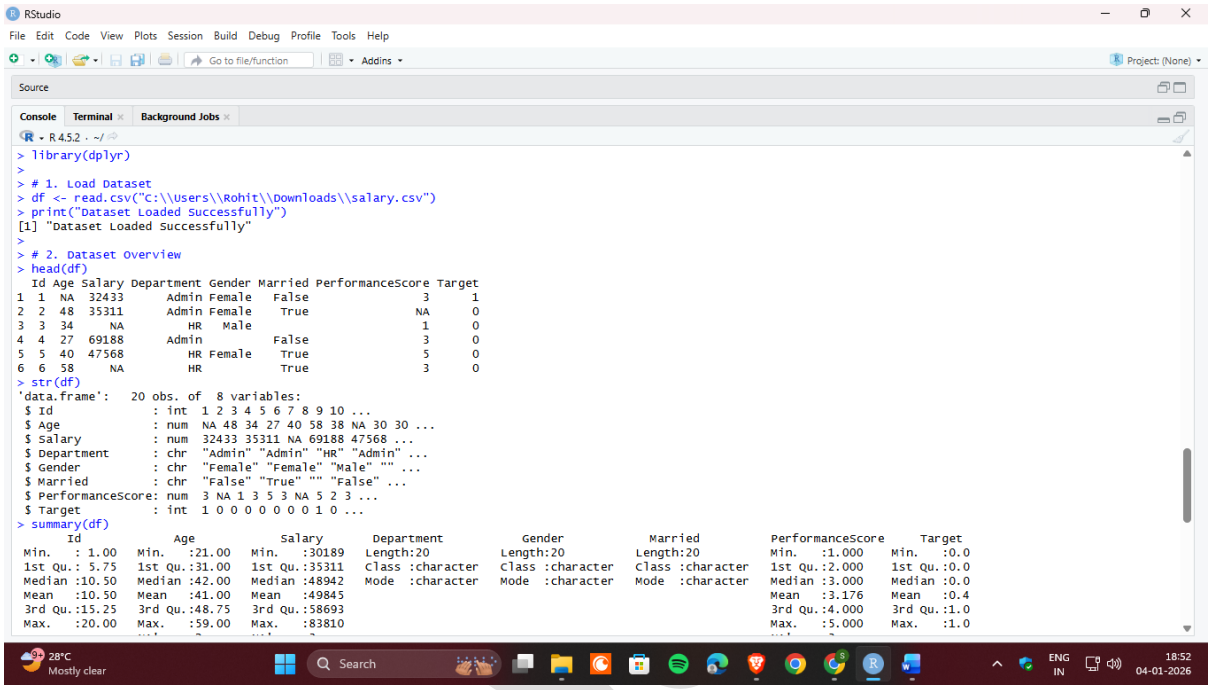
SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT :- R PROGRAMMING

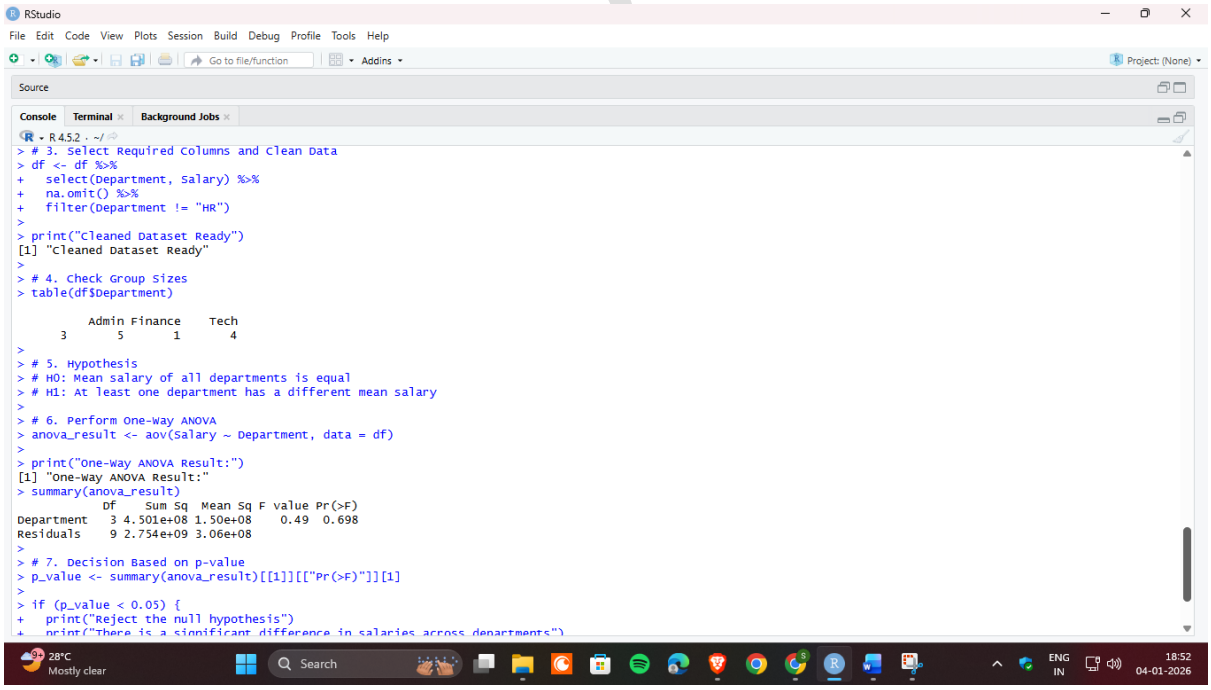
MODULE 2 – PRACTICAL 7

AIM: Performing one-way ANOVA using aov() (R).

OUTPUT:-



```
> library(dplyr)
>
> # 1. Load Dataset
> df <- read.csv("C:\\Users\\Rohit\\Downloads\\salary.csv")
> print("Dataset Loaded Successfully")
[1] "Dataset Loaded Successfully"
>
> # 2. Dataset Overview
> head(df)
  Id Age Salary Department Gender Married PerformanceScore Target
1  1  NA  32433      Admin Female   False                3      1
2  2  48  35311      Admin Female    True               NA      0
3  3  34     NA         HR   Male                       1      0
4  4  27  69188      Admin          False                3      0
5  5  40  47568         HR Female    True                5      0
6  6  58     NA         HR           True                3      0
> str(df)
'data.frame':   20 obs. of  8 variables:
 $ Id              : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Age             : num  NA 48 34 27 40 58 38 NA 30 30 ...
 $ Salary          : num  32433 35311 NA 69188 47568 ...
 $ Department      : chr  "Admin" "Admin" "HR" "Admin" ...
 $ Gender          : chr  "Female" "Female" "Male" "" ...
 $ Married         : chr  "False" "True" "" "False" ...
 $ PerformanceScore: num  3 NA 1 3 5 3 NA 5 2 3 ...
 $ Target          : int  1 0 0 0 0 0 0 0 1 0 ...
> summary(df)
       Id              Age            Salary       Department          Gender            Married          PerformanceScore     Target
 Min.   : 1.00   Min.   :21.00   Min.   :30189   Length:20          Length:20          Length:20          Min.   :1.000    Min.   :0.0
 1st Qu.: 5.75   1st Qu.:31.00   1st Qu.:35311   Class :character   Class :character   Class :character   1st Qu.:2.000    1st Qu.:0.0
 Median :10.50   Median :42.00   Median :48942   Mode  :character   Mode  :character   Mode  :character   Median :3.000    Median :0.0
 Mean   :10.50   Mean   :41.00   Mean   :49845                                                         Mean   :3.176    Mean   :0.4
 3rd Qu.:15.25   3rd Qu.:48.75   3rd Qu.:58693                                                         3rd Qu.:4.000    3rd Qu.:1.0
 Max.   :20.00   Max.   :59.00   Max.   :83810                                                         Max.   :5.000    Max.   :1.0
```



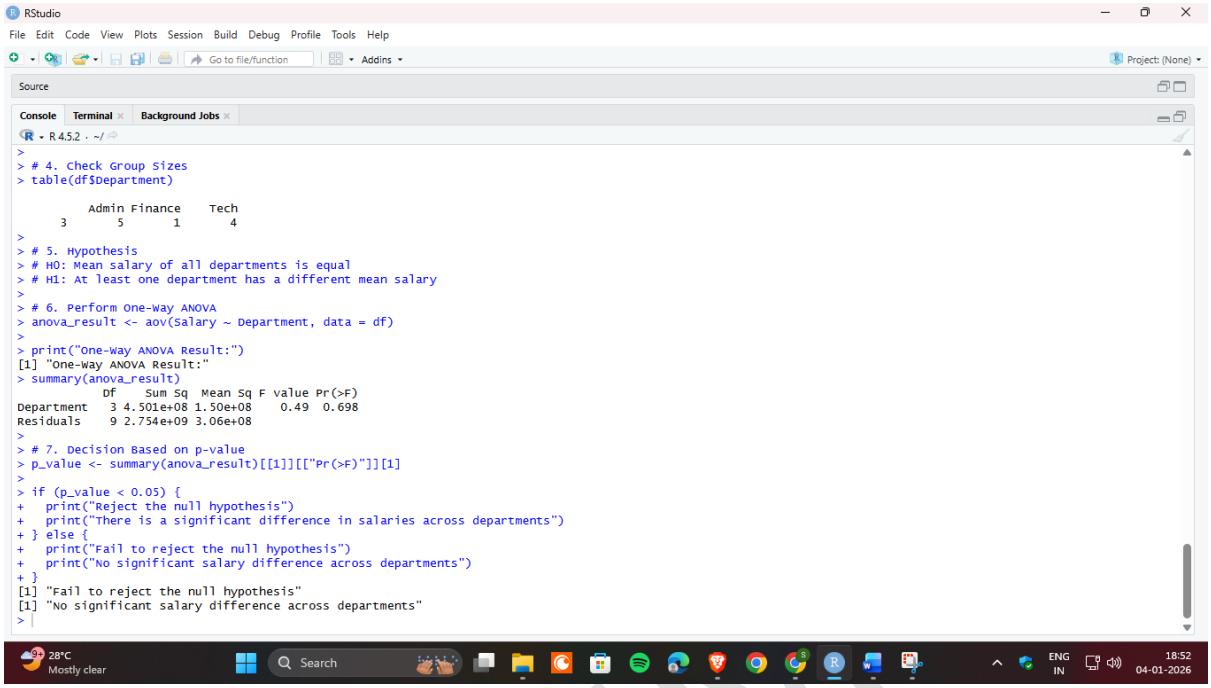
```
> # 3. Select Required Columns and Clean Data
> df <- df %>%
+   select(Department, Salary) %>%
+   na.omit() %>%
+   filter(Department != "HR")
>
> print("Cleaned Dataset Ready")
[1] "Cleaned Dataset Ready"
>
> # 4. Check Group Sizes
> table(df$Department)

         Admin Finance    Tech
     3       5       1       4
>
> # 5. Hypothesis
> # H0: Mean salary of all departments is equal
> # H1: At least one department has a different mean salary
>
> # 6. Perform One-Way ANOVA
> anova_result <- aov(Salary ~ Department, data = df)
>
> print("One-Way ANOVA Result:")
[1] "One-Way ANOVA Result:"
> summary(anova_result)
            Df    Sum Sq  Mean Sq F value Pr(>F)
Department   3 4.501e+08 1.50e+08    0.49  0.698
Residuals    9 2.754e+09 3.06e+08
>
> # 7. Decision Based on p-value
> p_value <- summary(anova_result)[[1]][["Pr(>F)"]][1]
>
> if (p_value < 0.05) {
+   print("Reject the null hypothesis")
+   print("There is a significant difference in salaries across departments")
```

SUMEET YADAV
S124

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

SUBJECT :- R PROGRAMMING



SUMEET YADAV

S124