

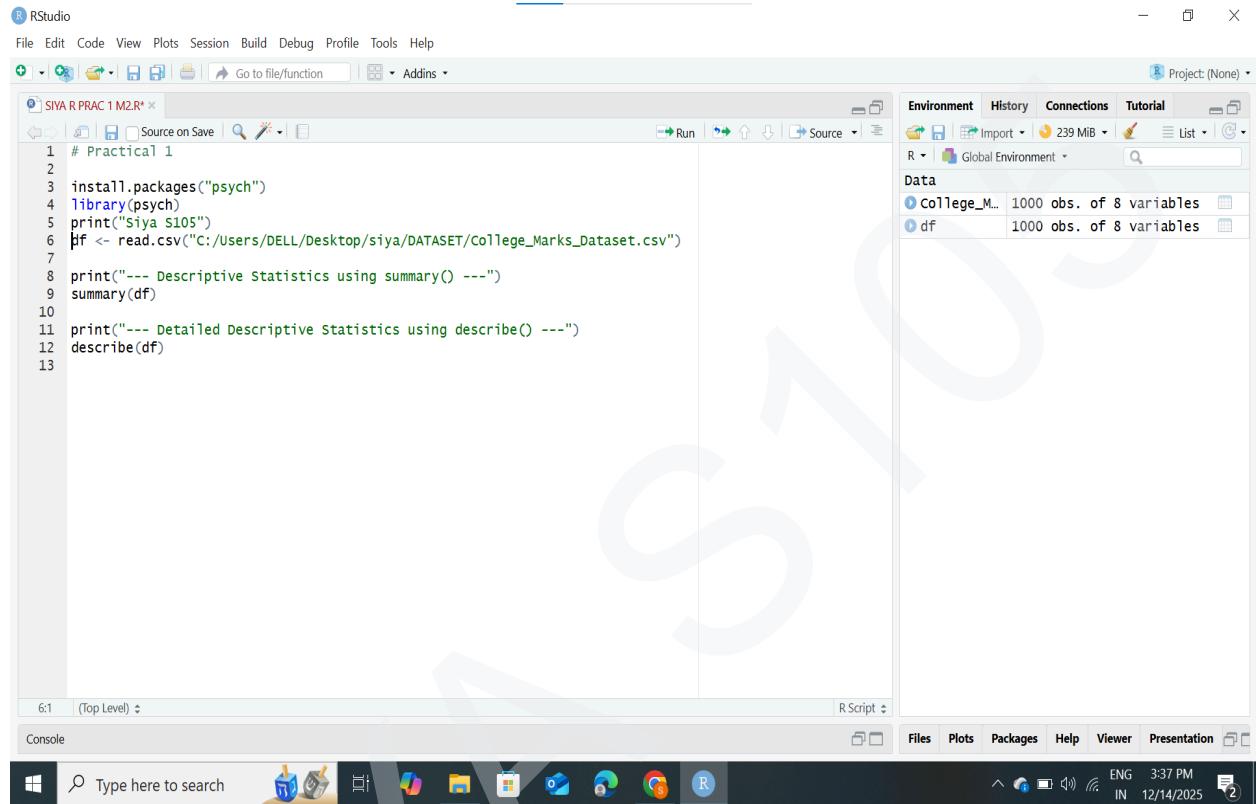
MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 1 TO 6

Module 2: Practical 1 to 6

AIM: 1. Generating descriptive statistics using `summary()` or `describe()` (R)



The screenshot shows the RStudio interface with the following details:

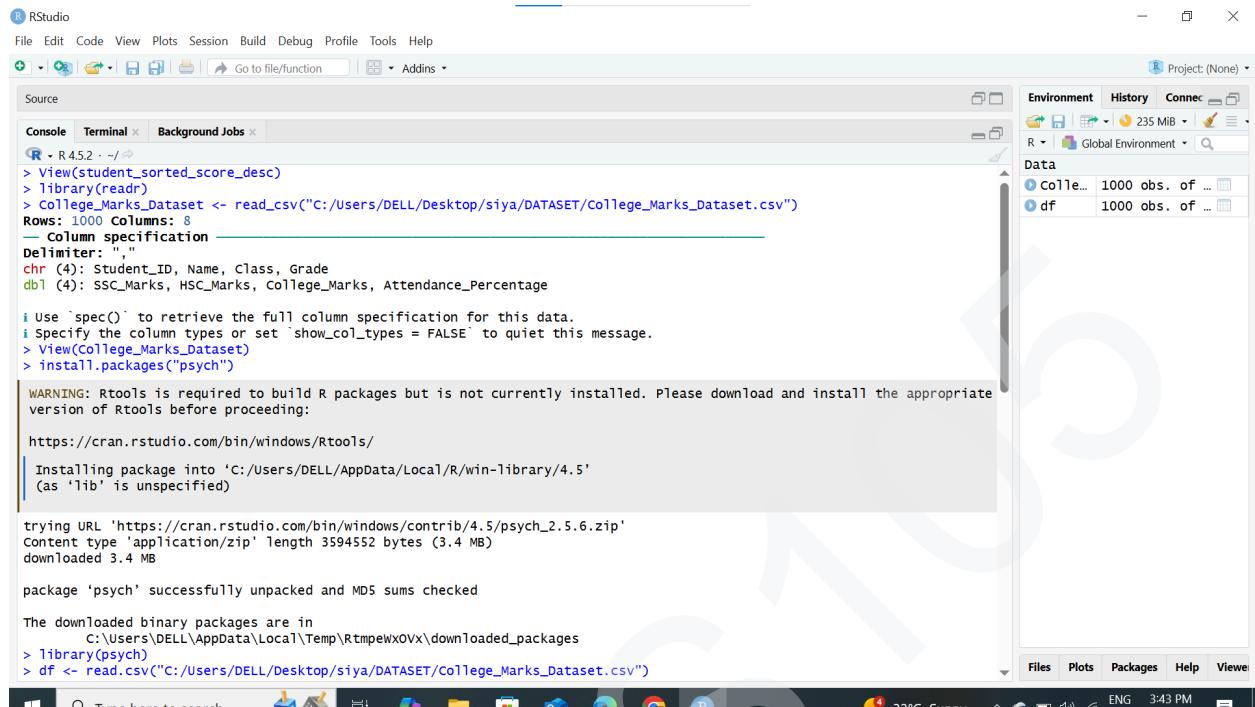
- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Source on Save, Run, Source, Environment, History, Connections, Tutorial, Project (None).
- Script Editor:** SIYA R PRAC 1 M2.R* contains the following R code:

```
1 # Practical 1
2
3 install.packages("psych")
4 library(psych)
5 print("SIYA S105")
6 df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/College_Marks_Dataset.csv")
7
8 print("--- Descriptive Statistics using summary() ---")
9 summary(df)
10
11 print("--- Detailed Descriptive Statistics using describe() ---")
12 describe(df)
13
```
- Data View:** Shows two datasets: College_M... (1000 obs. of 8 variables) and df (1000 obs. of 8 variables).
- Console:** Shows the command `source("SIYA R PRAC 1 M2.R")`.
- Bottom Bar:** Includes the Windows taskbar with various pinned icons like File Explorer, Edge, and R, along with system status icons.

MVLU COLLEGE

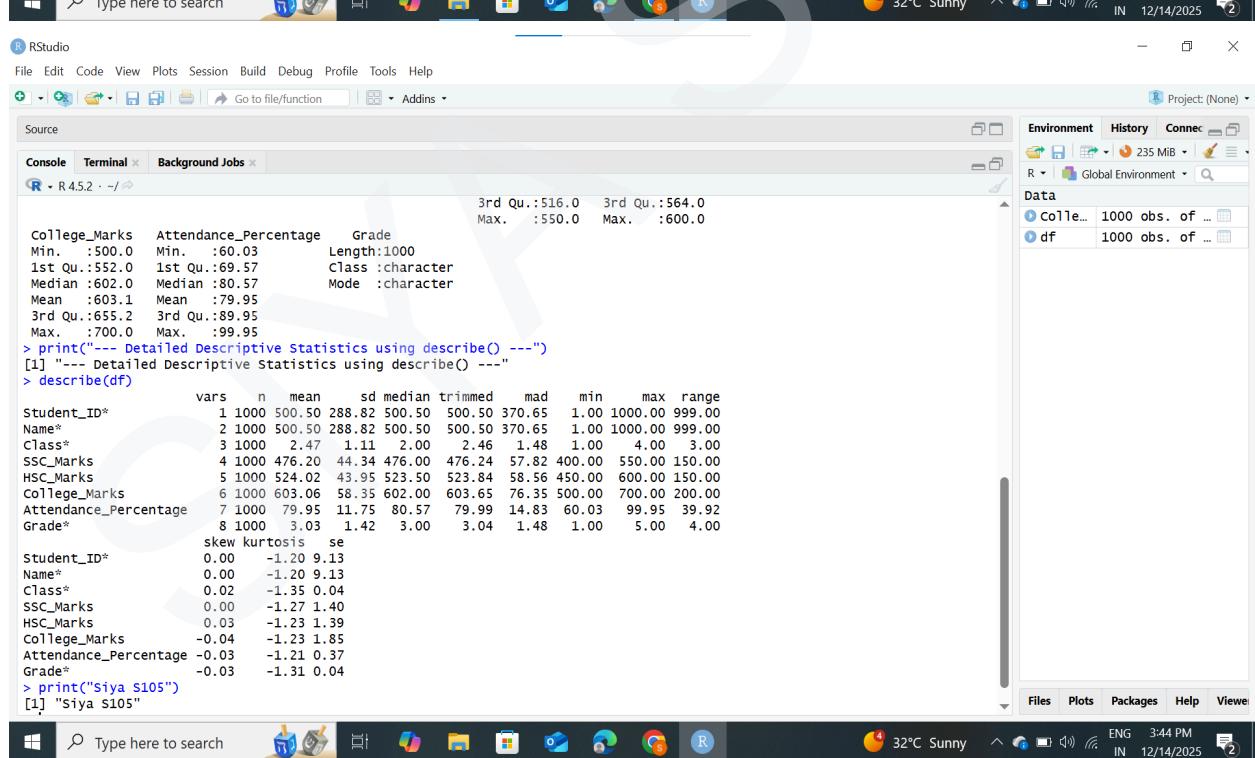
R PROGRAMMING

PRACTICAL NO. 1 TO 6



The screenshot shows the RStudio interface with the following details:

- Console Tab:** Displays R code and its output. The code includes loading a dataset, installing packages, and attempting to install Rtools. It also shows the download and unpacking of the psych package.
- Data View:** Shows two data frames: "College_Marks_Dataset" and "df", both containing 1000 observations.
- Environment View:** Shows the global environment with objects like "College_Marks_Dataset" and "df".
- Windows Taskbar:** Shows the Windows taskbar with the RStudio icon.



The screenshot shows the RStudio interface with the following details:

- Console Tab:** Displays R code and its output. The code includes printing descriptive statistics for the dataset using the `describe` function.
- Data View:** Shows two data frames: "College_Marks_Dataset" and "df", both containing 1000 observations.
- Environment View:** Shows the global environment with objects like "College_Marks_Dataset" and "df".
- Windows Taskbar:** Shows the Windows taskbar with the RStudio icon.

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 1 TO 6

AIM: 2. Generating frequency tables using table() or count() (R).

The screenshot shows the RStudio interface with the following details:

- File Menu:** File Edit Code View Plots Session Build Debug Profile Tools Help
- Source Editor:** SIYA R PRAC 1 M2.R* and SIYA R PRAC 2 M2.R* are open. The code in SIYA R PRAC 2 M2.R* is:

```
1 # Practical 2: Frequency Tables
2
3 install.packages("dplyr")
4 library(dplyr)
5 print("Siya S105")
6 df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/sales_data.csv")
7 colnames(df)
8
9 print("--- Frequency Table using table() ---")
10 category_counts <- table(df$Product_Category)
11 print(category_counts)
12
13 print("--- Frequency Table using count() ---")
14 category_df <- df %>% count(Product_Category)
15 print(category_df)
```

- Environment View:** Shows objects in the Global Environment:
 - category_df: 4 obs. of 2 variables
 - college_Marks...: 1000 obs. of 8 variables
 - df: 1000 obs. of 14 variables
 - sales_data: 1000 obs. of 14 variables
- Values View:** Shows values for category_counts and product_counts.
- Console:** Displays the command "R Script" and the status bar shows "16:1 (Top Level)".
- Taskbar:** Includes icons for File, Plots, Packages, Help, Viewer, and Presentation.
- System Tray:** Shows system information: 31°C Sunny, ENG 4:51 PM IN 12/14/2025.

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 1 TO 6

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ - Go to file/function Addins
Source
Console Terminal Background Jobs
R - R 4.5.2 - ~/>
> install.packages("dplyr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/DELL/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/dplyr_1.1.4.zip'
Content type 'application/zip' length 1593482 bytes (1.5 MB)
downloaded 1.5 MB

package 'dplyr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/Users/DELL/AppData/Local/Temp/Rtmpewx0vX/downloaded_packages
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':
  filter, lag

The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

> print("SIYA S105")
[1] "SIYA S105"

```

Environment History Connect

Data

- category 4 obs. of 2 v...
- College... 1000 obs. of ...
- df 1000 obs. of ...
- sales... 1000 obs. of ...

Values

- category 'table' int [1:...]
- product 'table' int[0 :...

Files Plots Packages Help Viewer

31°C Sunny ENG 4:52 PM IN 12/14/2025


```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ - Go to file/function Addins
Source
Console Terminal Background Jobs
R - R 4.5.2 - ~/>
< table of extent 0 >
> print("---- Frequency Table using count() ---")
[1] "---- Frequency Table using count() ---"
> product_df <- df %>% count(Product)
Error in `count()`:
! Must group by variables found in `.data`.
  x Column `Product` is not found.
Run rlang::last_trace() to see where the error occurred.

> colnames(df)
[1] "Product_ID"          "Sale_Date"           "Sales_Rep"          "Region"
[5] "Sales_Amount"         "Quantity_Sold"       "Product_Category"  "Unit_Cost"
[9] "Unit_Price"          "Customer_Type"       "Discount"          "Payment_Method"
[13] "Sales_Channel"        "Region_and_Sales_Rep"
> print("---- Frequency Table using table() ---")
[1] "---- Frequency Table using table() ---"
> category_counts <- table(df$Product_Category)
> print(category_counts)

Clothing Electronics Food Furniture
268      246    226    260
> print("---- Frequency Table using count() ---")
[1] "---- Frequency Table using count() ---"
> category_df <- df %>% count(Product_Category)
> print(category_df)

Product_Category n
1 Clothing 268
2 Electronics 246
3 Food 226
4 Furniture 260

```

Environment History Connect

Data

- category 4 obs. of 2 v...
- College... 1000 obs. of ...
- df 1000 obs. of ...
- sales... 1000 obs. of ...

Values

- category 'table' int [1:...]
- product 'table' int[0 :...

Files Plots Packages Help Viewer

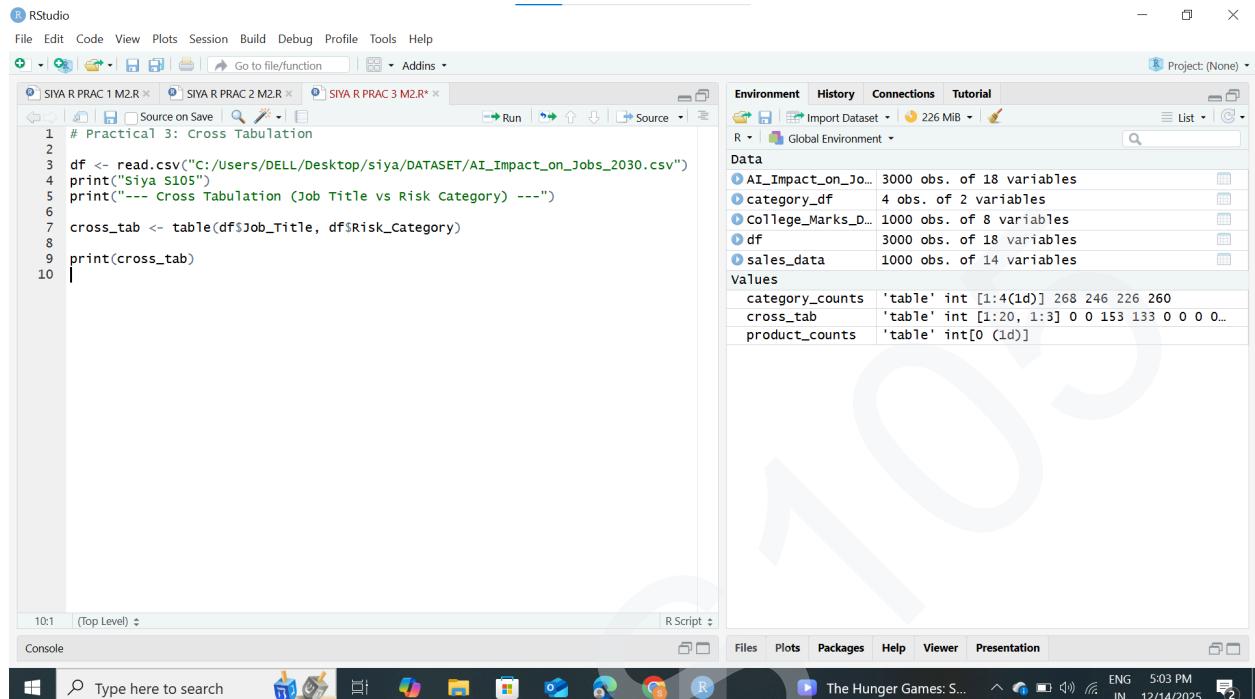
31°C Sunny ENG 4:52 PM IN 12/14/2025

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 1 TO 6

AIM: 3. Creating cross-tabulations and two-way tables using table() (R).



```

# Practical 3: Cross Tabulation
df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/AI_Impact_on_Jobs_2030.csv")
print("Siya S105")
print("--- Cross Tabulation (Job Title vs Risk Category) ---")
cross_tab <- table(df$Job_Title, df$Risk_Category)
print(cross_tab)

```

The screenshot shows the RStudio interface with the code above. The Environment pane displays the following objects:

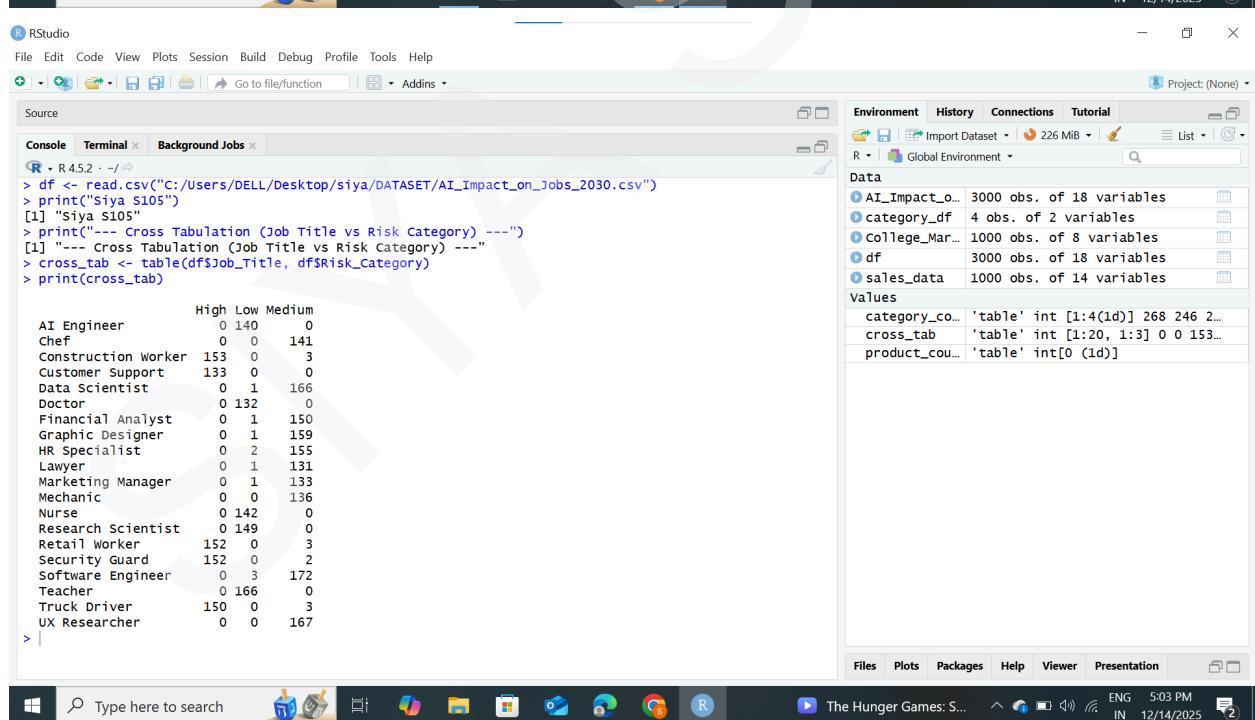
- AI_Impact_on_Jo... 3000 obs. of 18 variables
- category_df 4 obs. of 2 variables
- College_Marks_D... 1000 obs. of 8 variables
- df 3000 obs. of 18 variables
- sales_data 1000 obs. of 14 variables

The Values pane shows the structure of the `cross_tab` object:

```

category_counts 'table' int [1:4(1d)] 268 246 226 260
cross_tab 'table' int [1:20, 1:3] 0 0 153 133 0 0 0 ...
product_counts 'table' int[0 (1d)]

```

```

> df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/AI_Impact_on_Jobs_2030.csv")
> print("Siya S105")
[1] "Siya S105"
> print("--- Cross Tabulation (Job Title vs Risk Category) ---")
[1] "--- Cross Tabulation (Job Title vs Risk Category) ---"
> cross_tab <- table(df$Job_Title, df$Risk_Category)
> print(cross_tab)

      High Low Medium
AI Engineer      0 140     0
Chef              0   0 141
Construction Worker 153   0     3
Customer Support 133   0     0
Data Scientist    0   1 166
Doctor            0 132     0
Financial Analyst 0   1 150
Graphic Designer  0   1 159
HR Specialist     0   2 155
Lawyer             0   1 131
Marketing Manager 0   1 133
Mechanic           0   0 136
Nurse              0 142     0
Research Scientist 0 149     0
Retail Worker      152   0     3
Security Guard    152   0     2
Software Engineer 0   3 172
Teacher            0 166     0
Truck Driver      150   0     3
UX Researcher     0   0 167

```

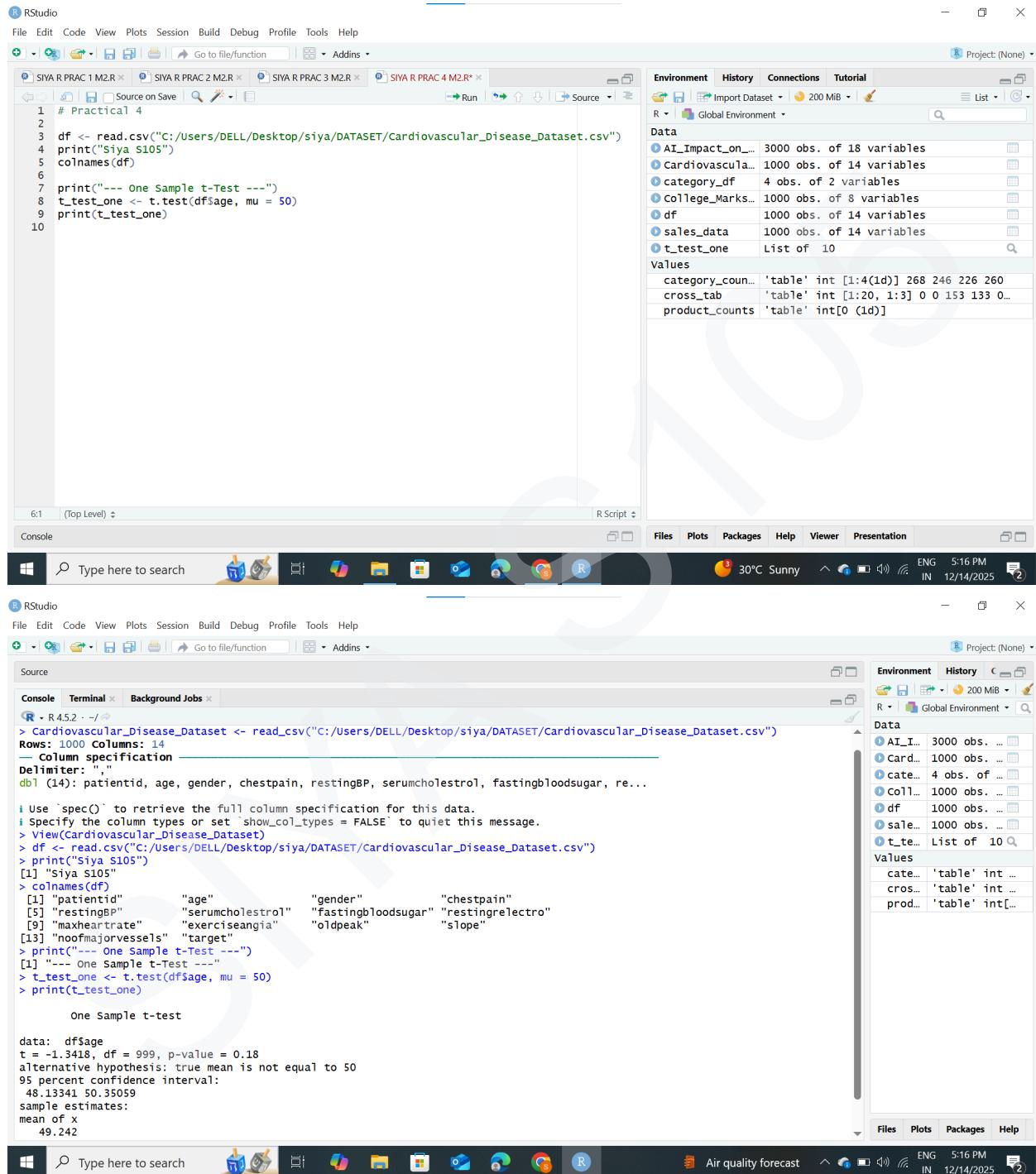
The screenshot shows the RStudio interface with the code above. The Environment pane displays the same objects as the first screenshot. The Values pane shows the structure of the `cross_tab` object again.

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 1 TO 6

AIM: 4. Performing one-sample t-tests using t.test) (R).



The screenshot shows two instances of RStudio. The top instance displays the R script and environment pane, while the bottom instance shows the console and environment panes.

R Script (Top Instance):

```

1 # Practical 4
2
3 df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/Cardiovascular_Disease_Dataset.csv")
4 print("Siya S105")
5 colnames(df)
6
7 print("---- One Sample t-Test ---")
8 t_test_one <- t.test(df$age, mu = 50)
9 print(t_test_one)
10

```

Environment (Top Instance):

- AI_Impact_on... 3000 obs. of 18 variables
- Cardiovascula... 1000 obs. of 14 variables
- category_df 4 obs. of 2 variables
- college_Marks... 1000 obs. of 8 variables
- df 1000 obs. of 14 variables
- sales_data 1000 obs. of 14 variables
- t_test_one List of 10

Values (Top Instance):

- category_count... 'table' int [1:4(1d)] 268 246 226 260
- cross_tab 'table' int [1:20, 1:3] 0 0 153 133 0...
- product_counts 'table' int[0 (1d)]

Console (Bottom Instance):

```

> Cardiovascular_Disease_Dataset <- read_csv("C:/Users/DELL/Desktop/siya/DATASET/Cardiovascular_Disease_Dataset.csv")
Rows: 1000 Columns: 14
Column specification:
Delimiter: ","
dbl (14): patientid, age, gender, chestpain, restingBP, serumcholesterol, fastingbloodsugar, re...
Use 'spec()' to retrieve the full column specification for this data.
Specify the column types or set `show_col_types = FALSE` to quiet this message.
> View(Cardiovascular_Disease_Dataset)
> df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/Cardiovascular_Disease_Dataset.csv")
> print("Siya S105")
[1] "Siya S105"
> colnames(df)
[1] "patientid"      "age"           "gender"         "chestpain"       "restingBP"       "serumcholesterol" "fastingbloodsugar" "restingrelectro" 
[5] "maxheartrate"   "exerciseangia" "oldpeak"        "slope"          "noofmajorvessels" "target"        
> print("---- One Sample t-Test ---")
[1] "---- One Sample t-Test ---"
> t_test_one <- t.test(df$age, mu = 50)
> print(t_test_one)

One Sample t-test

data: df$age
t = -1.3418, df = 999, p-value = 0.18
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
48.13341 50.35059
sample estimates:
mean of x
49.242

```

Environment (Bottom Instance):

- AI_I... 3000 obs. ...
- Card... 1000 obs. ...
- cate... 4 obs. of ...
- coi... 1000 obs. ...
- df 1000 obs. ...
- sale... 1000 obs. ...
- t_te... List of 10

Values (Bottom Instance):

- cate... 'table' int ...
- cros... 'table' int ...
- prod... 'table' int[...]

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 1 TO 6

AIM: 5. Performing independent two-sample t-tests using `t.test()` with grouping (R)

The screenshot shows the RStudio interface with the following details:

- File Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Source on Save, Go to file/function, Run, Source.
- Code Editor:**

```

1 # Practical 5
2
3 df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/Cleaned_Car_Price_Prediction.csv")
4 print("Siya S105")
5 colnames(df)
6 df$Price_Group <- ifelse(df$price > median(df$price), "High", "Low")
7
8 print("--- Independent Two-Sample t-Test ---")
9
10 t_test_two <- t.test(price ~ Price_Group, data = df)
11
12 print(t_test_two)
13

```
- Environment View:** Shows objects like AI_Impact_on..., Cardiovascul..., category_df, etc.
- Console View:** Shows the R session output for the practical code.
- Bottom Status Bar:** Shows system information: 30°C Sunny, ENG, 5:22 PM, IN, 12/14/2025.

The screenshot shows the RStudio interface with the following details:

- File Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Source on Save, Go to file/function, Run, Source.
- Code Editor:**

```

> Cleaned_Car_Price_Prediction <- read_csv("C:/Users/DELL/Desktop/siya/DATASET/Cleaned_Car_Price_Prediction.csv")
Rows: 199 Columns: 10
--- Column specification ---
Delimiter: ","
chr (5): brand, fuel_type, transmission, condition, model
dbl (5): car_id, year, engine_size, mileage, price

# Use `spec()` to retrieve the full column specification for this data.
# Specify the column types or set `show_col_types = FALSE` to quiet this message.
> View(Cleaned_Car_Price_Prediction)
> df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/Cleaned_Car_Price_Prediction.csv")
> print("Siya S105")
[1] "Siya S105"
> colnames(df)
[1] "car_id"      "brand"        "year"         "engine_size"   "fuel_type"    "transmission"  "mileage"      "condition"
[9] "price"       "model"
> df$Price_Group <- ifelse(df$price > median(df$price), "High", "Low")
> print("--- Independent Two-Sample t-Test ---")
[1] "--- Independent Two-Sample t-Test ---"
> t_test_two <- t.test(price ~ Price_Group, data = df)
> print(t_test_two)

Welch Two Sample t-test

data: price by Price_Group
t = 26.329, df = 196.91, p-value < 2.2e-16
alternative hypothesis: true difference in means between group High and group Low is not equal to 0
95 percent confidence interval:
45120.93 52427.50
sample estimates:
mean in group High mean in group Low
74560.68 25786.47

```
- Environment View:** Shows objects like AI_Impact_on..., Car..., cat..., etc.
- Console View:** Shows the R session output for the practical code.
- Bottom Status Bar:** Shows system information: 30°C Sunny, ENG, 5:23 PM, IN, 12/14/2025.

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 1 TO 6

AIM: 6. Performing paired t-tests using `t.test(paired=TRUE)`(R).



The screenshot shows the RStudio environment with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for file operations like Open, Save, Print, and Run.
- Source Editor:** Displays the R script for "Practical 6: Paired t-Test".
- Environment View:** Shows the global environment with objects like `df`, `t_test_paired`, and `category_df`.
- Console:** Shows the command prompt and search bar.
- Taskbar:** Includes icons for Files, Plots, Packages, Help, Viewer, and Presentation.
- System Tray:** Shows system status like ENG, 5:49 PM, IN, and date/time.

```
1 # Practical 6: Paired t-Test
2
3 df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/Cleaned_BMW_Sales_Data.csv")
4 print("Siya S105")
5 colnames(df)
6 print("--- Paired t-Test ---")
7
8 df$previous_year_sales <- df$sales_volume - runif(
9   nrow(df),
10   min = 50,
11   max = 200
12 )
13
14 t_test_paired <- t.test(
15   df$sales_volume,
16   df$previous_year_sales,
17   paired = TRUE
18 )
19
20 print(t_test_paired)
21 |
```

MVLU COLLEGE

R PROGRAMMING

PRACTICAL NO. 1 TO 6

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

Source

Console Terminal Background Jobs

```
R - R 4.5.2 - /<-->
> library(readr)
> Cleaned_BMW_Sales_Data <- read_csv("C:/Users/DELL/Desktop/siya/DATASET/Cleaned_BMW_Sales_Data.csv")
Rows: 99 Columns: 11
--- column specification ---
Delimiter: ","
chr (6): model, region, color, fuel_type, transmission, sales_classification
dbl (5): year, engine_size_1, mileage_km, price_usd, sales_volume

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
> View(Cleaned_BMW_Sales_Data)
> df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/Cleaned_BMW_Sales_Data.csv")
> print("Siya S105")
[1] "Siya S105"
> colnames(df)
[1] "model"           "year"            "region"          "color"           "fuel_type"
[6] "transmission"    "engine_size_1"   "mileage_km"      "price_usd"       "sales_volume"
[11] "sales_classification"
> print("--- Paired t-Test ---")
[1] "--- Paired t-Test ---"
> df$previous_year_sales <- df$sales_volume - runif(
+   nrow(df),
+   min = 50,
+   max = 200
+ )
> t_test_paired <- t.test(
+   df$sales_volume,
+   df$previous_year_sales,
+   paired = TRUE
+ )
> print(t_test_paired)
```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ - × Go to file/function | Addins ×
Source
Console Terminal × Background Jobs ×
R - R 4.5.2 - /
> print(siva\$dfs)
[1] "siva \$105"
> colnames(df)
[1] "model" "year" "region"
[6] "transmission" "engine_size_l" "mileage_km"
[11] "color" "price_usd" "fuel_type"
[16] "sales_classification" "sales_volume"
> print("--- Paired t-Test ---")
[1] "--- Paired t-Test ---"
> df\$previous_year_sales <- df\$sales_volume - runif(
+ nrow(df),
+ min = 50,
+ max = 200
+)
> t_test_paired <- t.test(
+ df\$sales_volume,
+ df\$previous_year_sales,
+ paired = TRUE
+)
> print(t_test_paired)

Paired t-test

data: df\$sales_volume and df\$previous_year_sales
t = 28.028, df = 98, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
116.2890 134.0108
sample estimates:
mean difference
125.1499
> |