

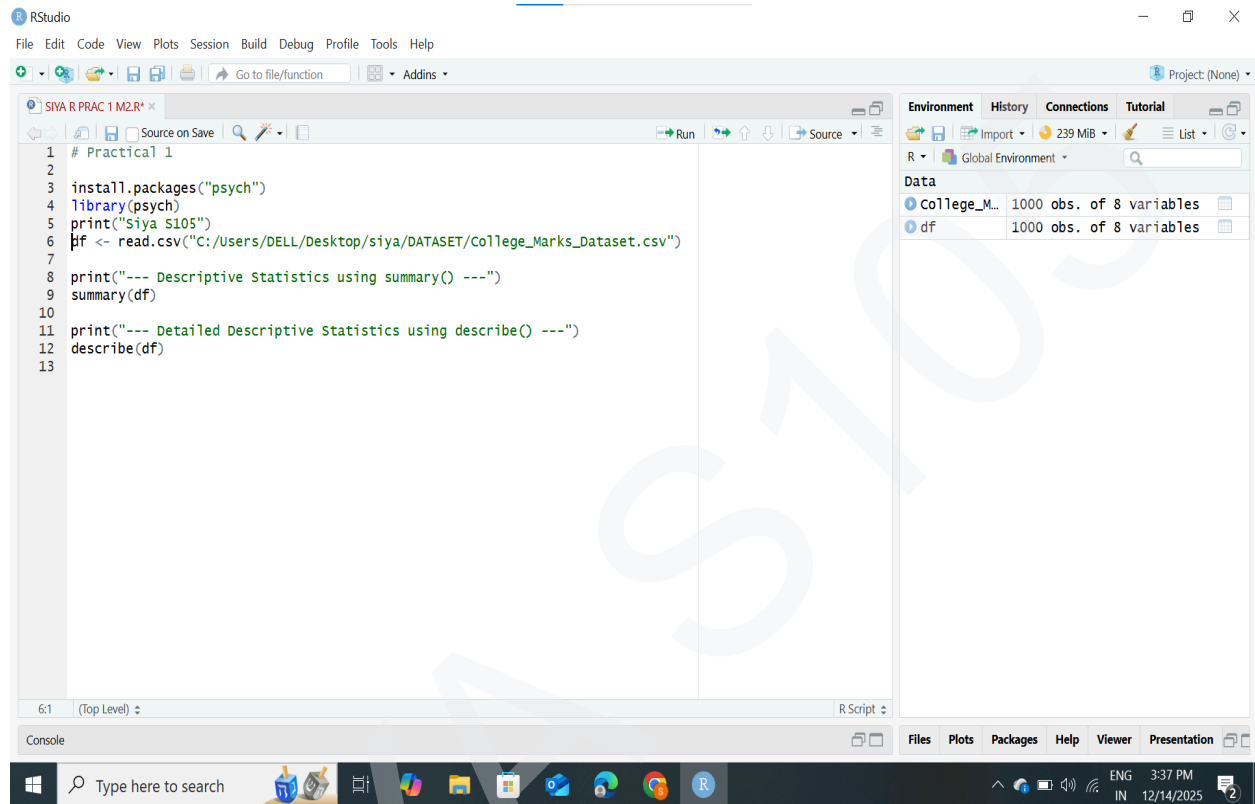
# MVLU COLLEGE

## R PROGRAMMING

### PRACTICAL NO. 1 TO 6

#### # Module 2: Practical 1 to 6

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



# 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)
Environment History Connec
R 235 MiB
Global Environment
Data
colle... 1000 obs. of ...
df 1000 obs. of ...
Files Plots Packages Help Viewe

Source
Console Terminal Background Jobs
R - R 4.5.2 - ~/
> View(student_sorted_score_desc)
> library(readr)
> college_Marks_Dataset <- read_csv("C:/Users/DELL/Desktop/siya/DATASET/college_Marks_Dataset.csv")
Rows: 1000 Columns: 8
Column specification
Delimiter: ","
chr (4): Student_ID, Name, Class, Grade
dbl (4): SSC_Marks, HSC_Marks, College_Marks, Attendance_Percentage
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(college_Marks_Dataset)
> install.packages("psych")

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/psych_2.5.6.zip'
Content type 'application/zip' length 3594552 bytes (3.4 MB)
downloaded 3.4 MB

package 'psych' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:/Users/DELL/AppData/Local/Temp/RtmpewxOVx/downloaded_packages
> library(psych)
> df <- read_csv("C:/Users/DELL/Desktop/siya/DATASET/College_Marks_Dataset.csv")
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Project: (None)
Environment History Connec
R 235 MiB
Global Environment
Data
colle... 1000 obs. of ...
df 1000 obs. of ...
Files Plots Packages Help Viewe

Source
Console Terminal Background Jobs
R - R 4.5.2 - ~/
3rd Qu.:516.0 3rd Qu.:564.0
Max.:550.0 Max.:600.0

College_Marks Attendance_Percentage Grade
Min.:500.0 Min.:60.03 Length:1000
1st Qu.:552.0 1st Qu.:69.57 Class:character
Median:602.0 Median:80.57 Mode:character
Mean:603.1 Mean:79.95
3rd Qu.:655.2 3rd Qu.:89.95
Max.:700.0 Max.:99.95
> print("--- Detailed Descriptive Statistics using describe() ---")
[1] "--- Detailed Descriptive Statistics using describe() ---"
> describe(df)
vars n mean sd median trimmed mad min max range
Student_ID* 1 1000 500.50 288.82 500.50 500.50 370.65 1.00 1000.00 999.00
Name* 2 1000 500.50 288.82 500.50 500.50 370.65 1.00 1000.00 999.00
Class* 3 1000 2.47 1.11 2.00 2.46 1.48 1.00 4.00 3.00
SSC_Marks 4 1000 476.20 44.34 476.00 476.24 57.82 400.00 550.00 150.00
HSC_Marks 5 1000 524.02 43.95 523.50 523.84 58.56 450.00 600.00 150.00
College_Marks 6 1000 603.06 58.35 602.00 603.65 76.35 500.00 700.00 200.00
Attendance_Percentage 7 1000 79.95 11.75 80.57 79.99 14.83 60.03 99.95 39.92
Grade* 8 1000 3.03 1.42 3.00 3.04 1.48 1.00 5.00 4.00
skew kurtosis se
Student_ID* 0.00 -1.20 9.13
Name* 0.00 -1.20 9.13
Class* 0.02 -1.35 0.04
SSC_Marks 0.00 -1.27 1.40
HSC_Marks 0.03 -1.23 1.39
College_Marks -0.04 -1.23 1.85
Attendance_Percentage -0.03 -1.21 0.37
Grade* -0.03 -1.31 0.04
> print("Siya S105")
[1] "Siya S105"
```

# MVLU COLLEGE

## R PROGRAMMING

### PRACTICAL NO. 1 TO 6

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

The screenshot shows the RStudio IDE with a script file named 'SIYA R PRAC 2 M2.R'. The code in the script is as follows:

```
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)
16
```

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

Data	
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

Below the data objects, the 'Values' section shows:

Values	
category_coun...	'table' int [1:4(1d)] 268 246 226 260
product_counts	'table' int[0 (1d)]

The console at the bottom is empty. The system tray at the bottom shows the date and time as 12/14/2025, 4:51 PM.

# MVLU COLLEGE

## R PROGRAMMING

### PRACTICAL NO. 1 TO 6

The screenshot shows the RStudio interface with the Console pane active. The user has run the command `install.packages("dplyr")`. The console output shows a warning about Rtools, the download of the dplyr package (1.5 MB), and its successful installation. It also lists the objects masked from the stats and base packages. The Environment pane on the right shows the global environment with variables like 'catego...', 'colleg...', 'df', and 'sales...'.

```
> 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\RtmpewxOVx\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"
```

The screenshot shows the RStudio interface with the Console pane active. The user has loaded the dplyr package and is working with a data frame 'df'. They have attempted to use `count()` but encountered an error because the 'Product' column was not found. The console output shows the error message and the user's subsequent commands to check the column names and create a frequency table using `table()`. The Environment pane on the right shows the global environment with variables like 'catego...', 'colleg...', 'df', and 'sales...'.

```
< 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
```

# MVLU COLLEGE

## R PROGRAMMING

### PRACTICAL NO. 1 TO 6

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

The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 # Practical 3: Cross Tabulation
2
3 df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/AI_Impact_on_Jobs_2030.csv")
4 print("Siya S105")
5 print("--- Cross Tabulation (Job Title vs Risk Category) ---")
6
7 cross_tab <- table(df$Job_Title, df$Risk_Category)
8
9 print(cross_tab)
10
```

The Environment pane on the right shows 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 following data:

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	0	0	0	0	0	0	0	0	0	0	0
product_counts	table' int[0 (1d)]																		

The screenshot shows the RStudio interface with the console output of the cross-tabulation table. The console output is as follows:

```
> 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)
```

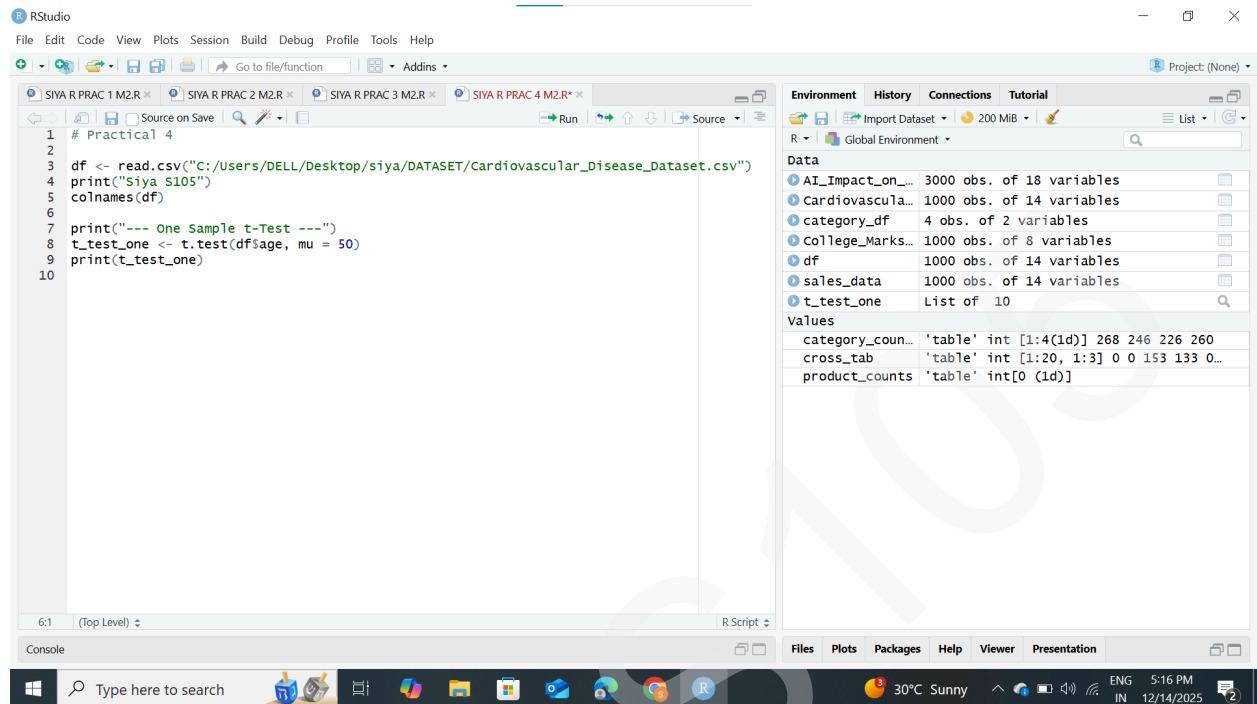
	High	Low	Medium
AI Engineer	0	140	0
Chef	0	0	141
Construction Worker	153	0	3
Customer Support	133	0	0
Data Scientist	0	1	166
Doctor	0	132	0
Financial Analyst	0	1	150
Graphic Designer	0	1	159
HR Specialist	0	2	155
Lawyer	0	1	131
Marketing Manager	0	1	133
Mechanic	0	0	136
Nurse	0	142	0
Research Scientist	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

# MVLU COLLEGE

## R PROGRAMMING

### PRACTICAL NO. 1 TO 6

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



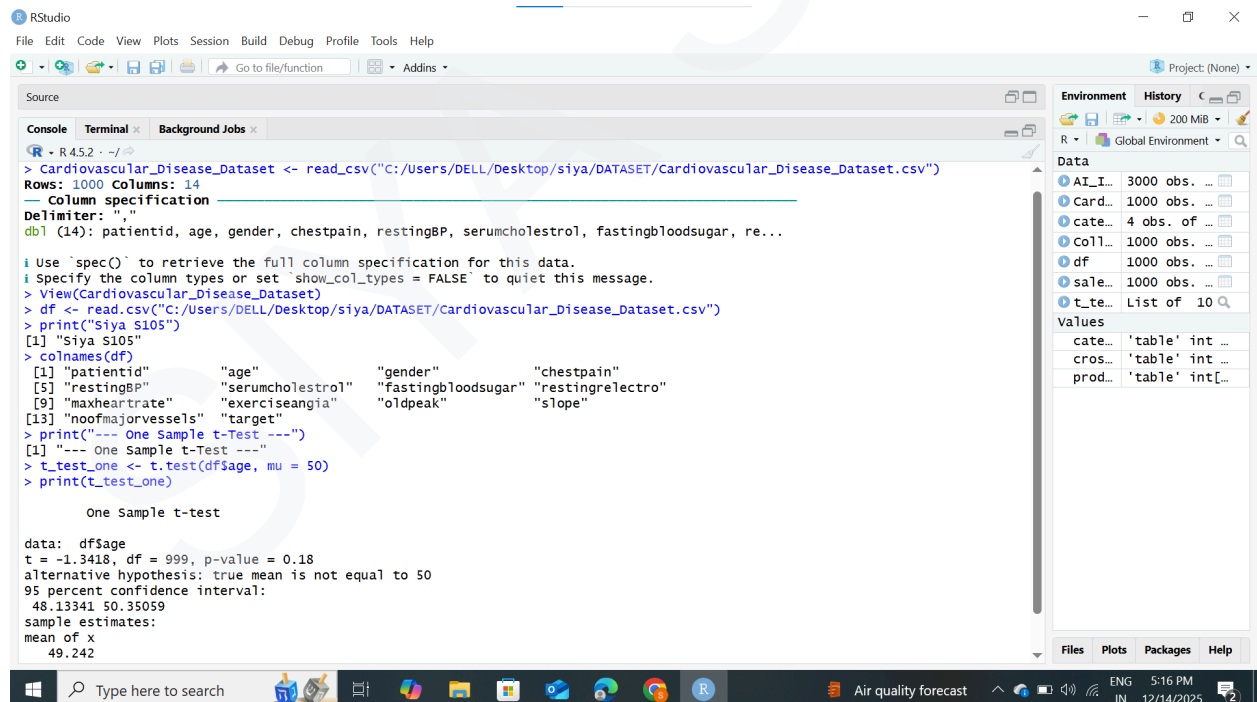
```
1 # Practical 4
2
3 df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/Cardiovascular_Disease_Dataset.csv")
4 print("Siya S105")
5 colnames(df)
6
7 print("--- One Sample t-Test ---")
8 t_test_one <- t.test(df$age, mu = 50)
9 print(t_test_one)
10
```

Environment

Object	Class	Attributes
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

Object	Class	Attributes
category_coun...	'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)]	



```
> 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...

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(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"
[5] "restingBP" "serumcholesterol" "fastingbloodsugar" "restingelectro"
[9] "maxheartrate" "exerciseargia" "oldpeak" "slope"
[13] "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

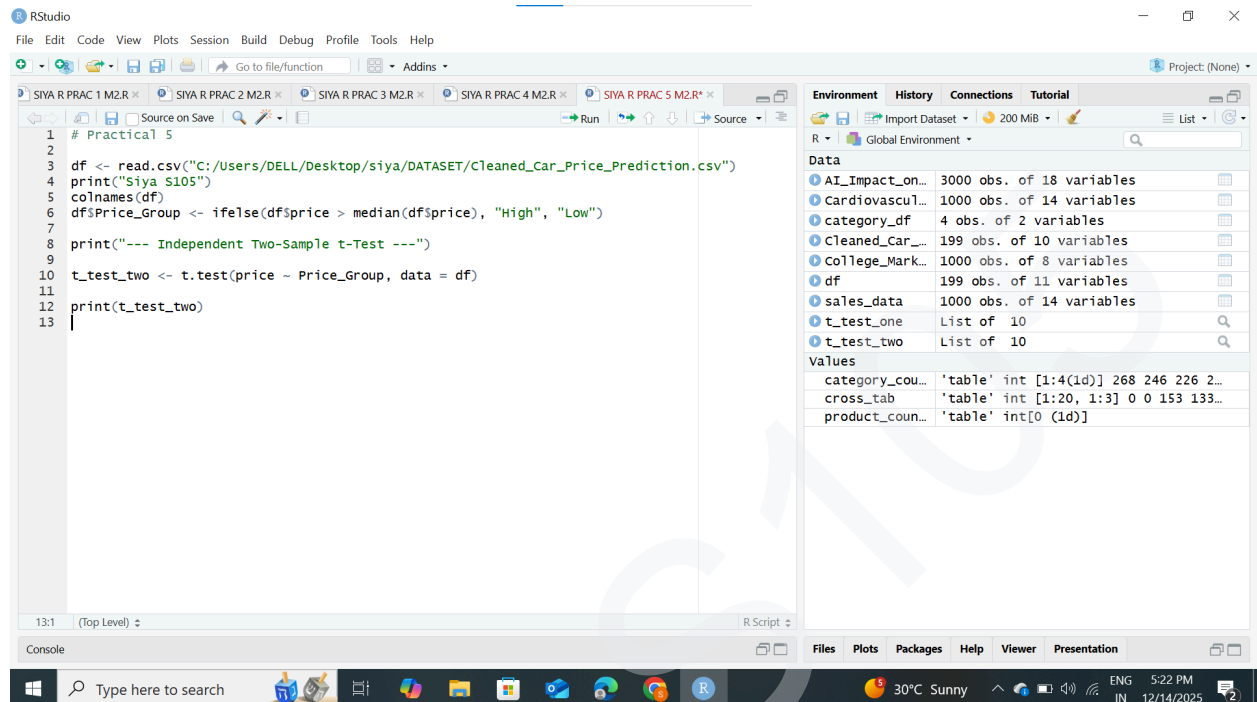
data: df$age
t = -1.3418, df = 999, p-value = 0.18
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
 48.13341 50.35059
sample estimates:
mean of x
 49.242
```

# MVLU COLLEGE

## R PROGRAMMING

### PRACTICAL NO. 1 TO 6

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



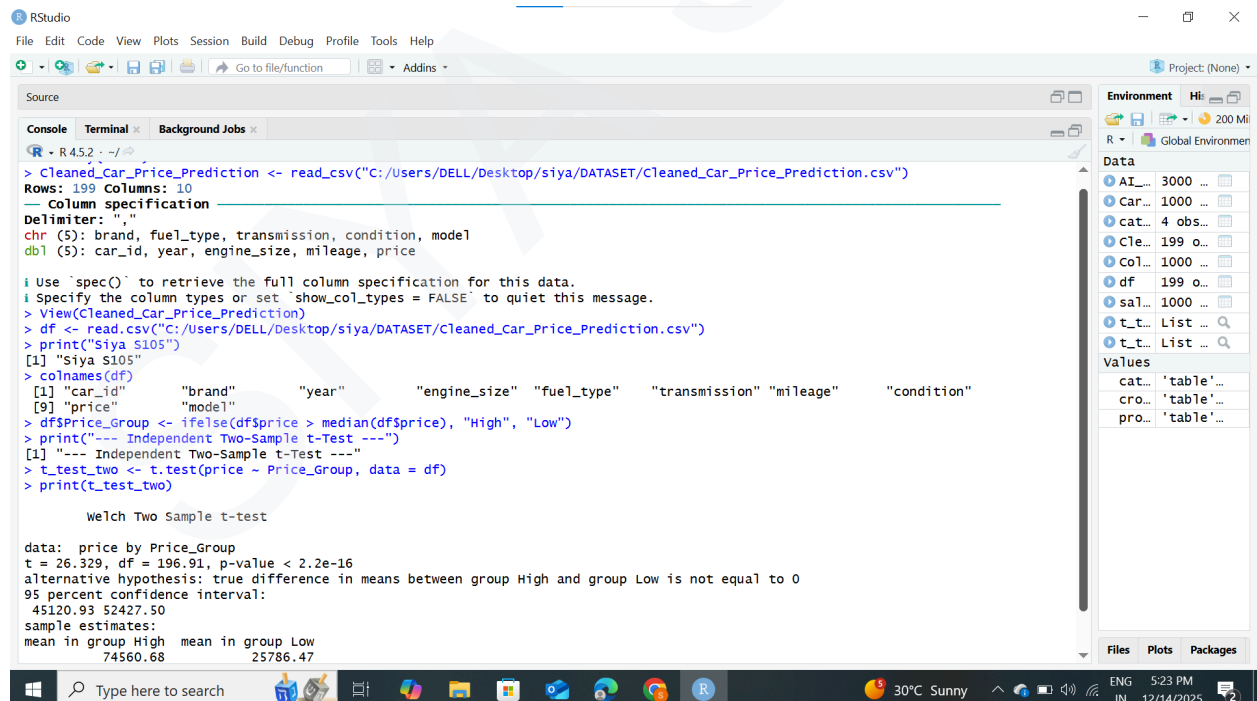
```
1 # Practical 5
2
3 df <- read.csv("C:/Users/DELL/Desktop/siya/DATASET/Cleaned_Car_Price_Prediction.csv")
4 print("Siya S105")
5 colnames(df)
6 df$Price_Group <- ifelse(df$price > median(df$price), "High", "Low")
7
8 print("--- Independent Two-Sample t-Test ---")
9
10 t_test_two <- t.test(price ~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

Environment

Object	Class	Attributes
AI_Impact_on...	3000 obs. of 18 variables	
Cardiovascul...	1000 obs. of 14 variables	
category_df	4 obs. of 2 variables	
Cleaned_Car_...	199 obs. of 10 variables	
College_Mark...	1000 obs. of 8 variables	
df	199 obs. of 11 variables	
sales_data	1000 obs. of 14 variables	
t_test_one	List of 10	
t_test_two	List of 10	

Values

Object	Class	Attributes
category_cou...	'table' int [1:4(id)]	268 246 226 2...
cross_tab	'table' int [1:20, 1:3]	0 0 153 133...
product_coun...	'table' int[0 (id)]	



```
> 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

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_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
```

# 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 interface with a script editor on the left and the Environment pane on the right. The script is titled "# Practical 6: Paired t-Test" and contains the following code:

```
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
```

The Environment pane on the right shows the following objects:

Object	Details
AI_Impact_on_...	3000 obs. of 18 variables
Cardiovascula...	1000 obs. of 14 variables
category_df	4 obs. of 2 variables
Cleaned_BMW_S...	99 obs. of 11 variables
Cleaned_Car_P...	199 obs. of 10 variables
college_Marks...	1000 obs. of 8 variables
df	99 obs. of 12 variables
sales_data	1000 obs. of 14 variables
t_test_one	List of 10
t_test_paired	List of 10
t_test_two	List of 10

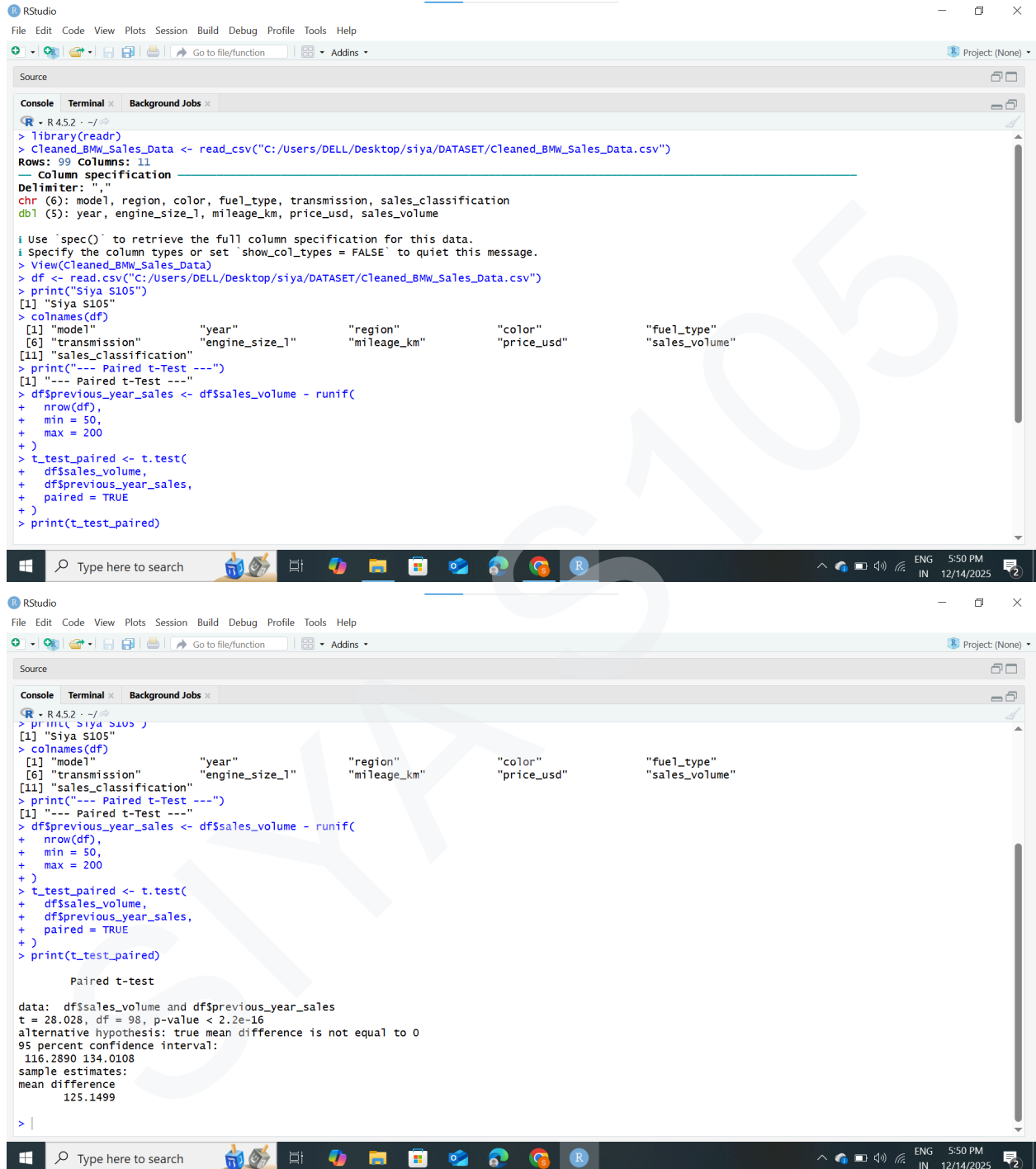
The console at the bottom shows the output of the script, including the print statement "Siya S105" and the results of the paired t-test.



# MVLU COLLEGE

## R PROGRAMMING

### PRACTICAL NO. 1 TO 6



The screenshot displays the RStudio interface with the following code and output:

```
> 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_l, mileage_km, price_usd, sales_volume

# 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_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_l" "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)
```

**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