

## MVLU COLLEGE

### PRACTICAL NO :- 07

**AIM :-** Selecting and dropping variables using select() in R. import dataset.

#### CODE :-

```
library(dplyr)

marks <- read.csv("C:/Users/itlab/Downloads/S100/Cleaned_Data_Science_Student_Marks.csv")

print("---- Original Dataset (First 3 rows) ----")
print(head(marks, 3))

# 2. SELECTING VARIABLES (Keeping Columns)
#method 1
# Scenario: We want only student_id, python_marks, and english_marks
selected_cols <- marks %>%
  select(student_id, python_marks, english_marks)

print("---- Selected Specific Columns ----")
print(head(selected_cols, 3))

#methods 2
# Scenario: Select columns from sql_marks to python_marks
range_cols <- marks %>%
  select(sql_marks:python_marks)
print("---- Selected Range of Columns (sql_marks to python_marks) ----")
print(head(range_cols, 3))

#method 3
# Scenario: Select all columns starting with "p" (python_marks, power_bi_marks)
starts_with_p <- marks %>%
  select(starts_with("p"))

print("---- Selected columns starting with 'p' ----")
print(head(starts_with_p, 3))

# 3. DROPPING VARIABLES (Removing Columns)
#Method A
# Scenario: Remove the 'location' column
dropped_one <- marks %>%
  select(-location)
print("---- Dataset with 'location' dropped ----")
print(names(dropped_one))

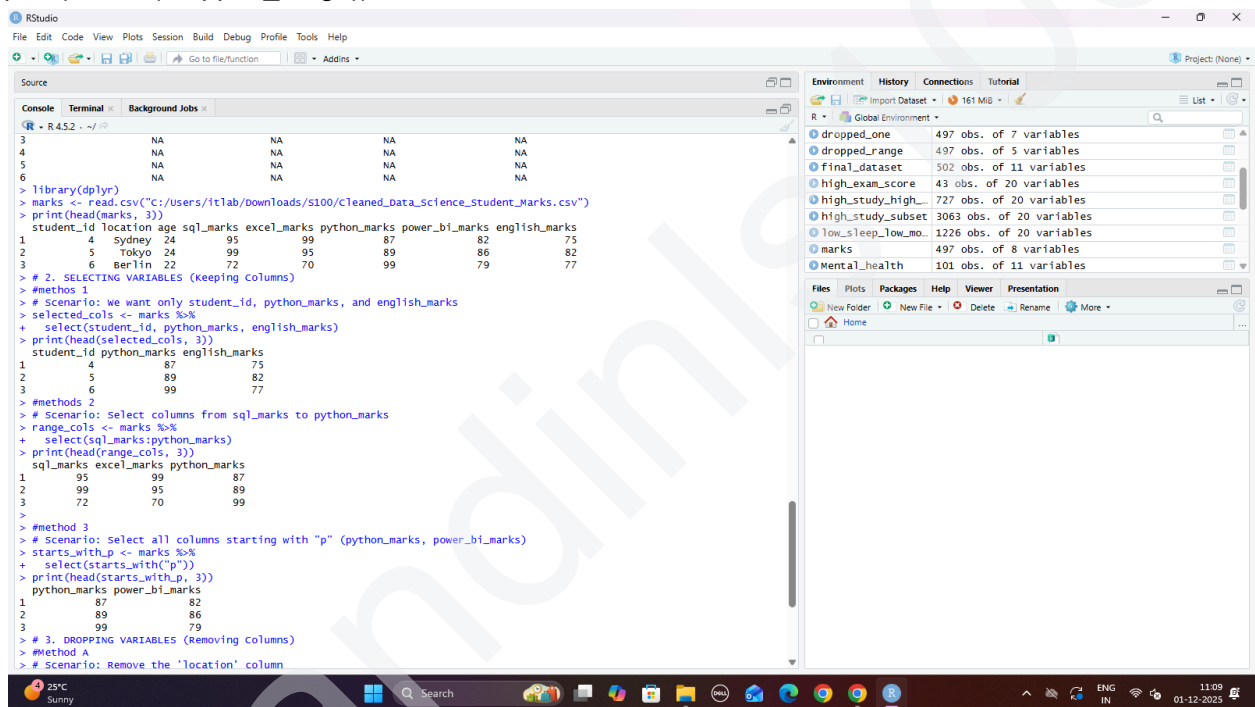
#Method B
```

## MVLU COLLEGE

```
# Scenario: Remove 'excel_marks' and 'power_bi_marks'
dropped_multiple <- marks %>%
  select(-excel_marks, -power_bi_marks)
print("--- Dataset with 'excel_marks' and 'power_bi_marks' dropped ---")
print(names(dropped_multiple))
```

### #Method C

```
# Scenario: Remove all columns from 'sql_marks' to 'python_marks'
dropped_range <- marks %>%
  select(-(sql_marks:python_marks))
print("--- Dataset with columns 'sql_marks' to 'python_marks' dropped ---")
print(names(dropped_range))
```



# MVLU COLLEGE

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for data manipulation using dplyr. The code includes comments for different methods and scenarios, such as selecting columns, dropping variables, and removing specific columns.
- Console:** Shows the output of the R code, including data frames and variable names.
- Environment:** Lists the objects in the R environment, such as 'dropped\_one', 'dropped\_range', 'final\_dataset', 'high\_exam\_score', 'high\_study\_high', 'high\_study\_subset', 'low\_sleep\_low\_mo', 'marks', and 'Mental\_health'.
- Files:** Shows the file explorer with a 'Home' directory.
- Taskbar:** Displays the Windows taskbar with various application icons and system tray information.

```
> # Scenario: Select columns from sql_marks to python_marks
> range_cols <- marks %>%
+   select(sql_marks:python_marks)
> print(head(range_cols, 3))
  sql_marks excel_marks python_marks
1         95         99         87
2         99         99         89
3         72         70         99

> # Method 3
> # Scenario: Select all columns starting with "p" (python_marks, power_bi_marks)
> starts_with_p <- marks %>%
+   select(starts_with("p"))
> print(head(starts_with_p, 3))
  python_marks power_bi_marks
1           87           82
2           89           86
3           99           79

> # 3. DROPPING VARIABLES (Removing columns)
> # Method A
> # Scenario: Remove the 'location' column
> dropped_one <- marks %>%
+   select(-location)
> print(names(dropped_one))
[1] "student_id" "age" "sql_marks" "excel_marks" "python_marks" "power_bi_marks"
[7] "english_marks"

> # Method B
> # Scenario: Remove 'excel_marks' and 'power_bi_marks'
> dropped_multiple <- marks %>%
+   select(-excel_marks, -power_bi_marks)
> print(names(dropped_multiple))
[1] "student_id" "location" "age" "sql_marks" "python_marks" "english_marks"

> # Method C
> # Scenario: Remove all columns from 'sql_marks' to 'python_marks'
> dropped_range <- marks %>%
+   select(-(sql_marks:python_marks))
> print(names(dropped_range))
[1] "student_id" "location" "age" "power_bi_marks" "english_marks"
```