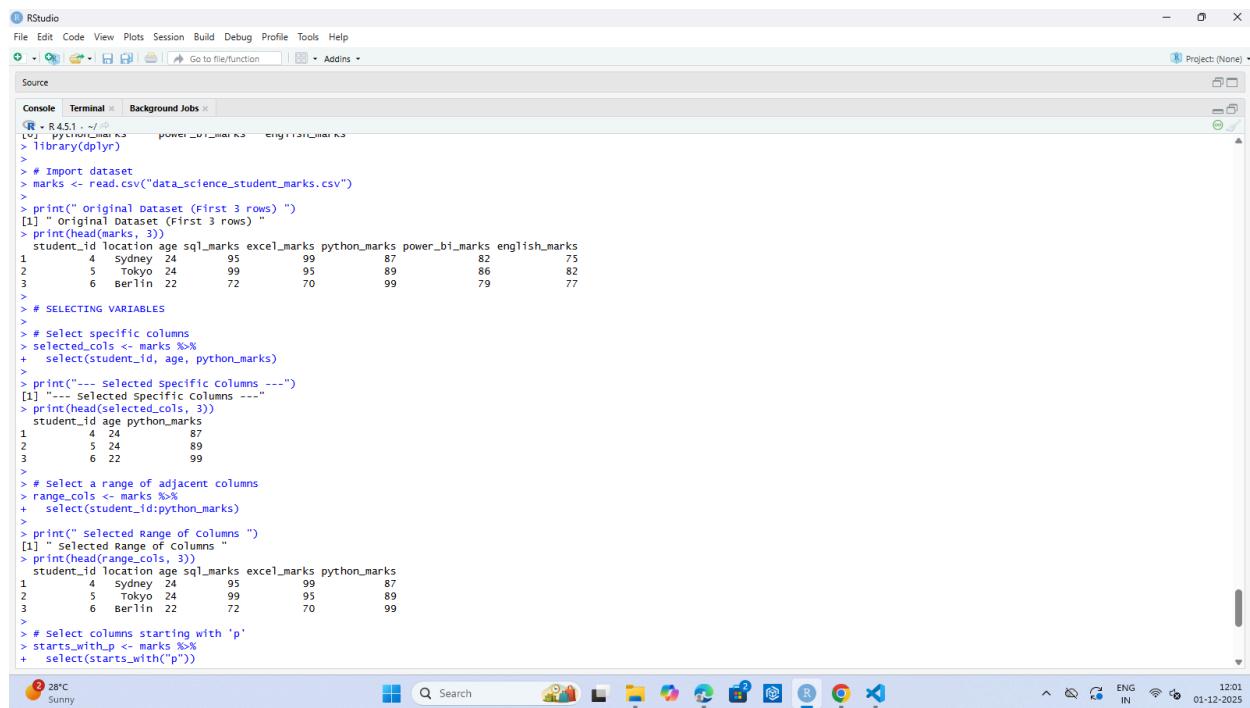


Practical No 7

Aim : Selecting and dropping variables using select() in R. import dataset.

Output :



The screenshot shows the RStudio interface with the following R code in the console:

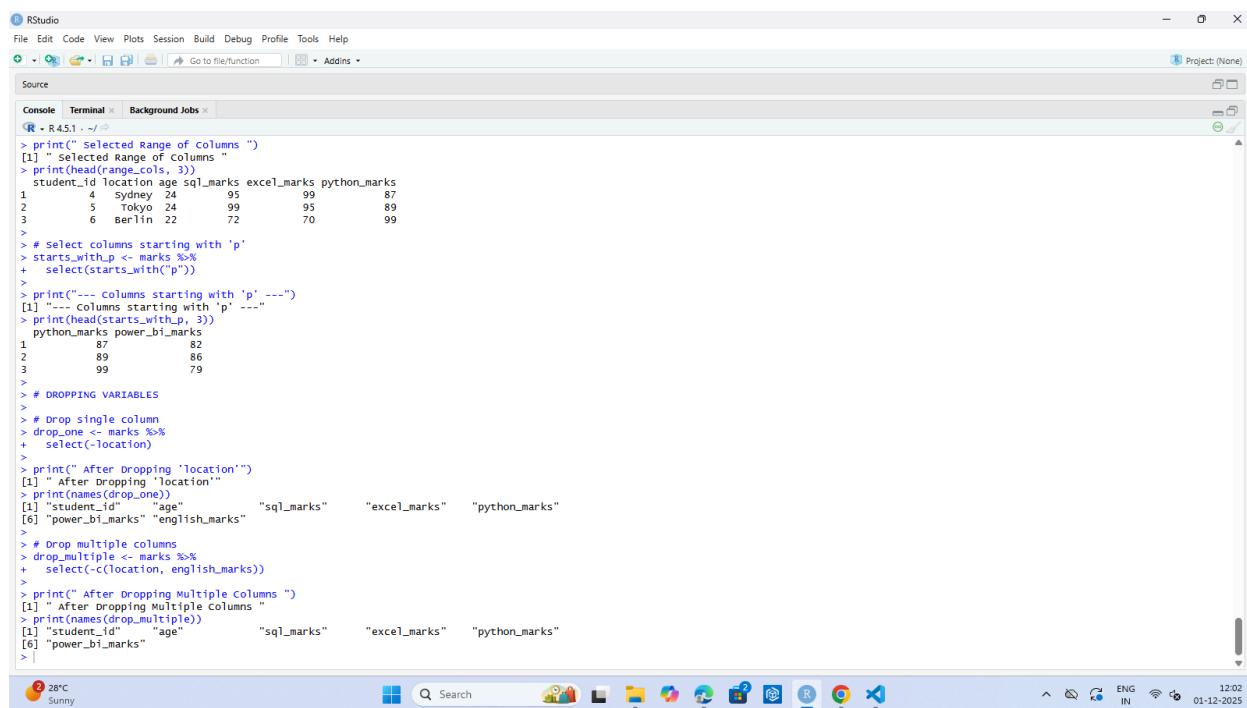
```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
> # Import dataset
> marks <- read.csv("data_science_student_marks.csv")
> print(" original dataset (First 3 rows) ")
[1] " original dataset (First 3 rows) "
> print(head(marks, 3))
#> # SELECTING VARIABLES
#> # select specific columns
#> selected_cols <- marks %>%
#>   select(student_id, age, python_marks)
#>
#> print("--- selected specific columns ---")
[1] "--- Selected Specific Columns ---"
> print(head(selected_cols, 3))
student_id age python_marks
1        4    24          87
2        5    24          89
3        6    22          99
#> # select a range of adjacent columns
#> range_cols <- marks %>%
#>   select(student_id:python_marks)
#>
#> print(" selected range of columns ")
[1] " Selected Range of Column "
> print(head(range_cols, 3))
student_id location age sql_marks excel_marks python_marks
1        4    Sydney 24      95       99          87
2        5    Tokyo   24      99       95          89
3        6    Berlin  22      72       70          99
#> # select columns starting with 'p'
#> starts_with_p <- marks %>%
#>   select(starts_with("p")))

```

The code imports a dataset named "data_science_student_marks.csv", prints the first three rows, and then performs three operations: selecting specific columns (student_id, age, python_marks), selecting a range of adjacent columns (student_id to python_marks), and selecting columns starting with 'p'.

SHETH L.U.J. AND SIR M.V. COLLEGE OF ARTS SCIENCE AND COMMERCE

SUBJECT : R Programming



The screenshot shows the RStudio interface with the following R code in the Console tab:

```
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 4.5.1 - ~/r
> print(" Selected Range of columns ")
[1] " Selected Range of columns "
> print(head(range_cols))
  Student_id location age sql_marks excel_marks python_marks
1           1    Sydney   24      95       99        87
2           5     Tokyo   24      99       95        89
3          6    Berlin   22      72       70        99
>
> # select columns starting with 'p'
> starts_with_p <- marks %>%
+   select(starts_with("p"))
>
> print("--- columns starting with 'p' ---")
[1] "--- columns starting with 'p' ---"
> print(head(starts_with_p, 3))
  python_marks power_bi_marks
1            87             82
2            89             86
3            99             79
>
> # DROPPING VARIABLES
>
> # drop single column
> drop_one <- marks %>%
+   select(-location)
>
> print(" After Dropping 'location'")
[1] " After Dropping 'location'"
> print(names(drop_one))
[1] "student_id" "age"           "sql_marks"      "excel_marks"   "python_marks"
[6] "power_bi_marks" "english_marks"
>
> # drop multiple columns
> drop_multiple <- marks %>%
+   select(-c(location, english_marks))
>
> print(" After Dropping Multiple columns ")
[1] " After Dropping Multiple columns "
> print(names(drop_multiple))
[1] "student_id" "age"           "sql_marks"      "excel_marks"   "python_marks"
[6] "power_bi_marks"
> |
```

The code demonstrates various R functions for data manipulation, including printing column ranges, selecting specific columns, dropping variables, and handling multiple columns.