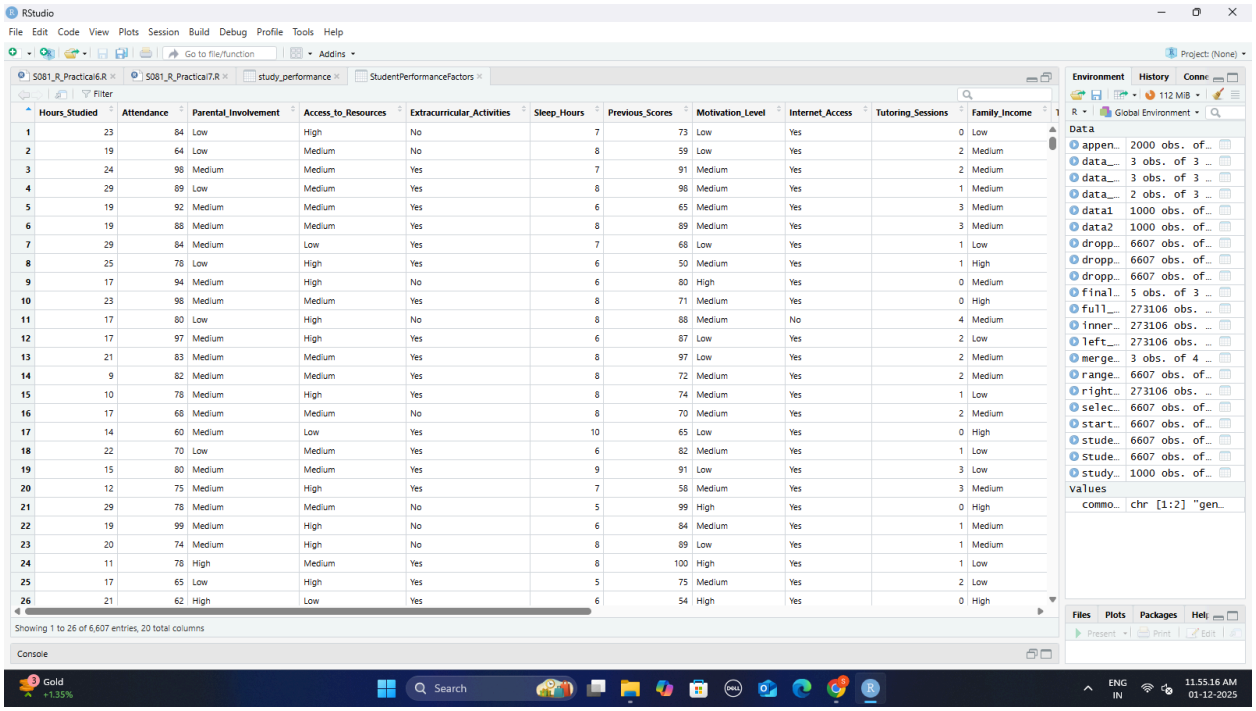
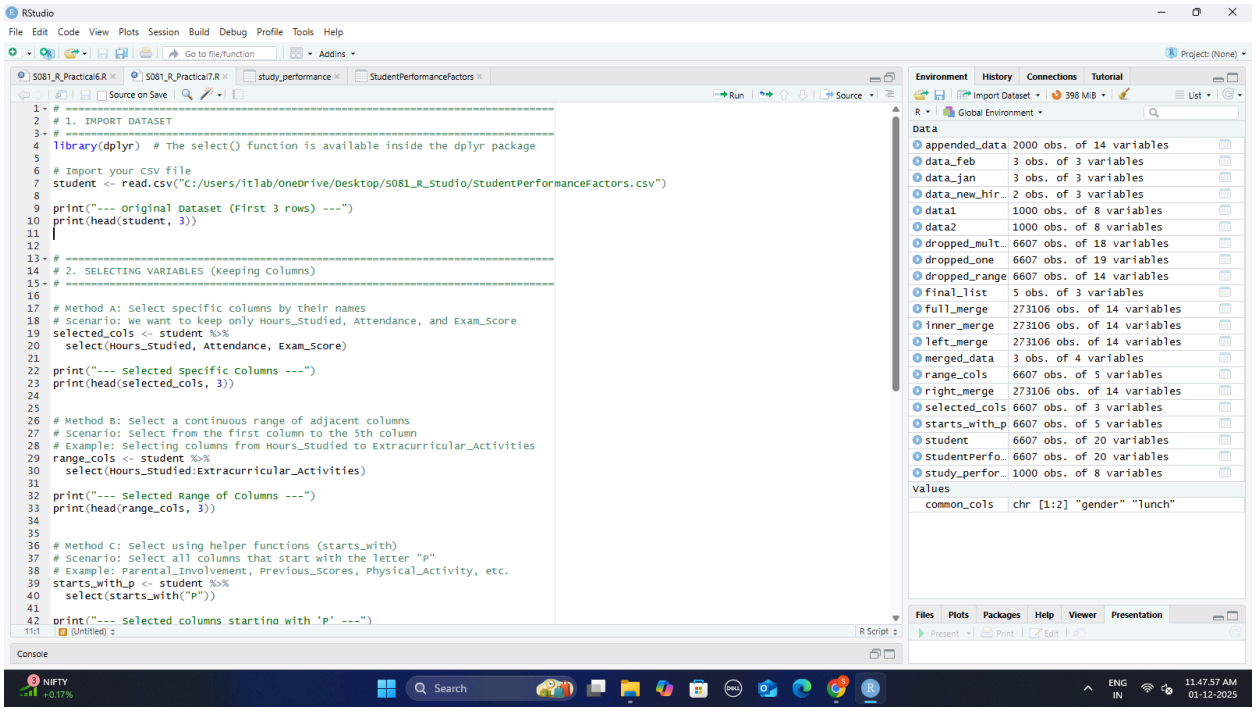
# Sheth L.U.J. & Sir M.V. College

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

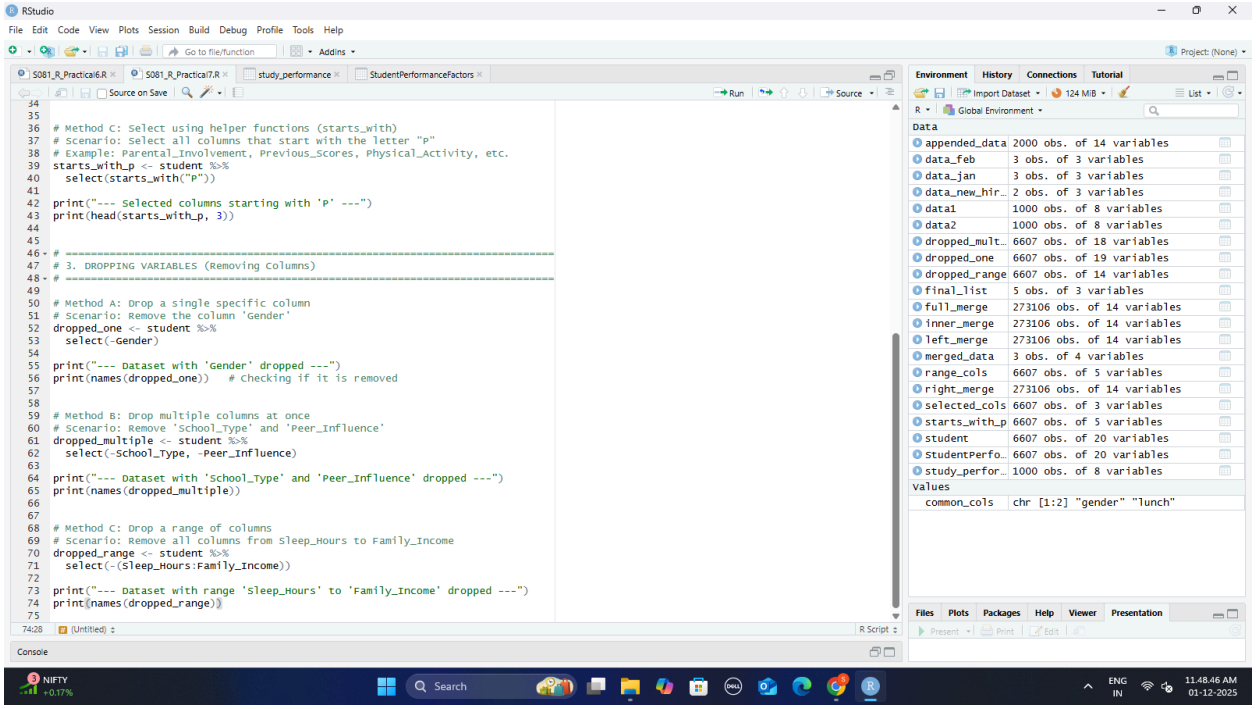


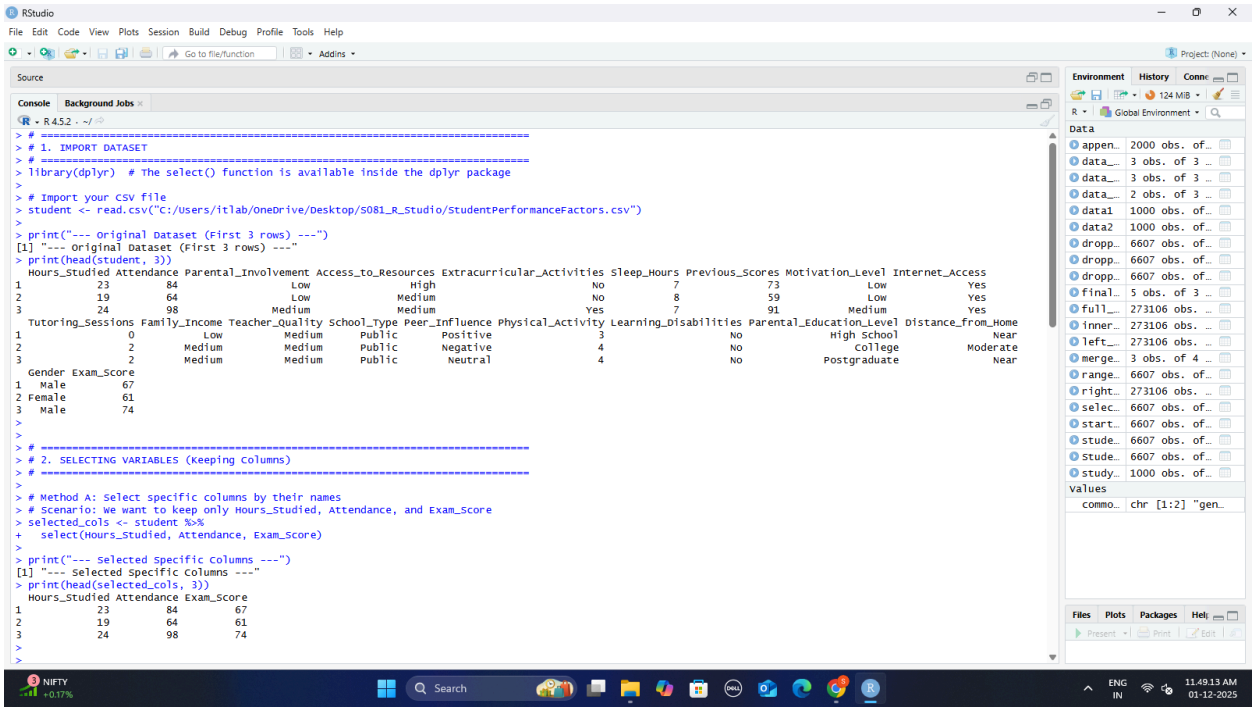


**Name :- Priya Gupta**

**Roll no. :- S081**

# Sheth L.U.J. & Sir M.V. College

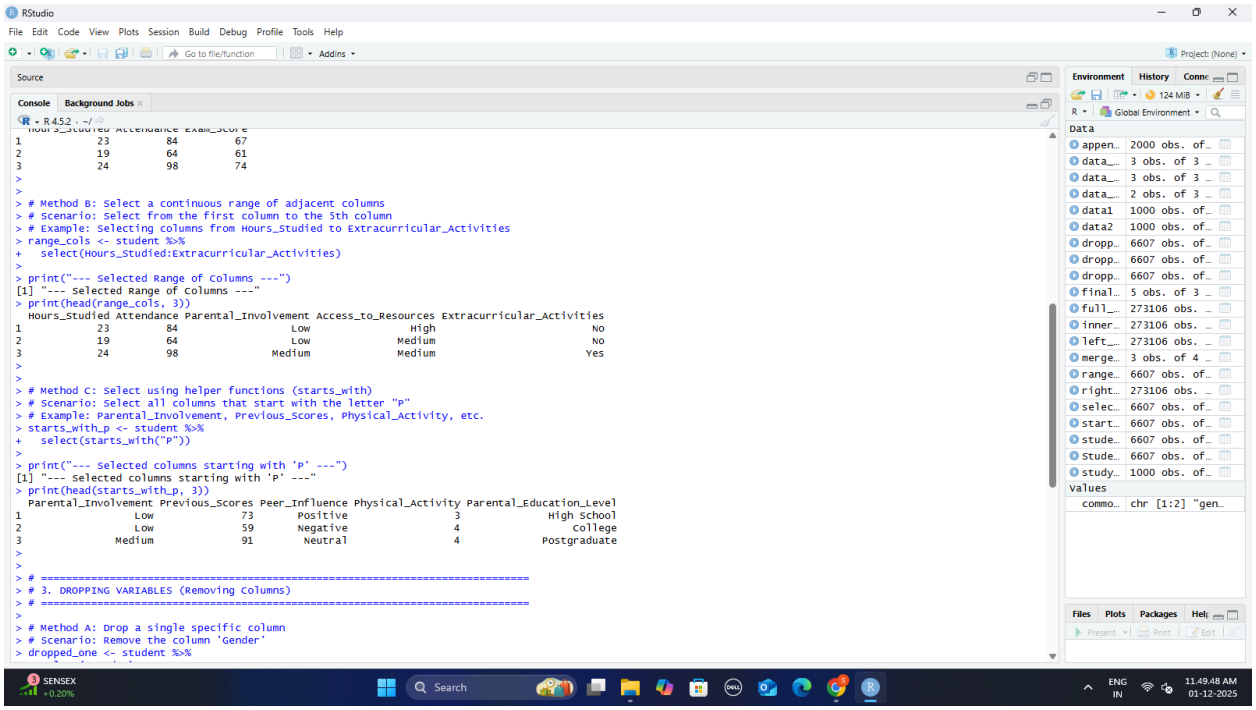


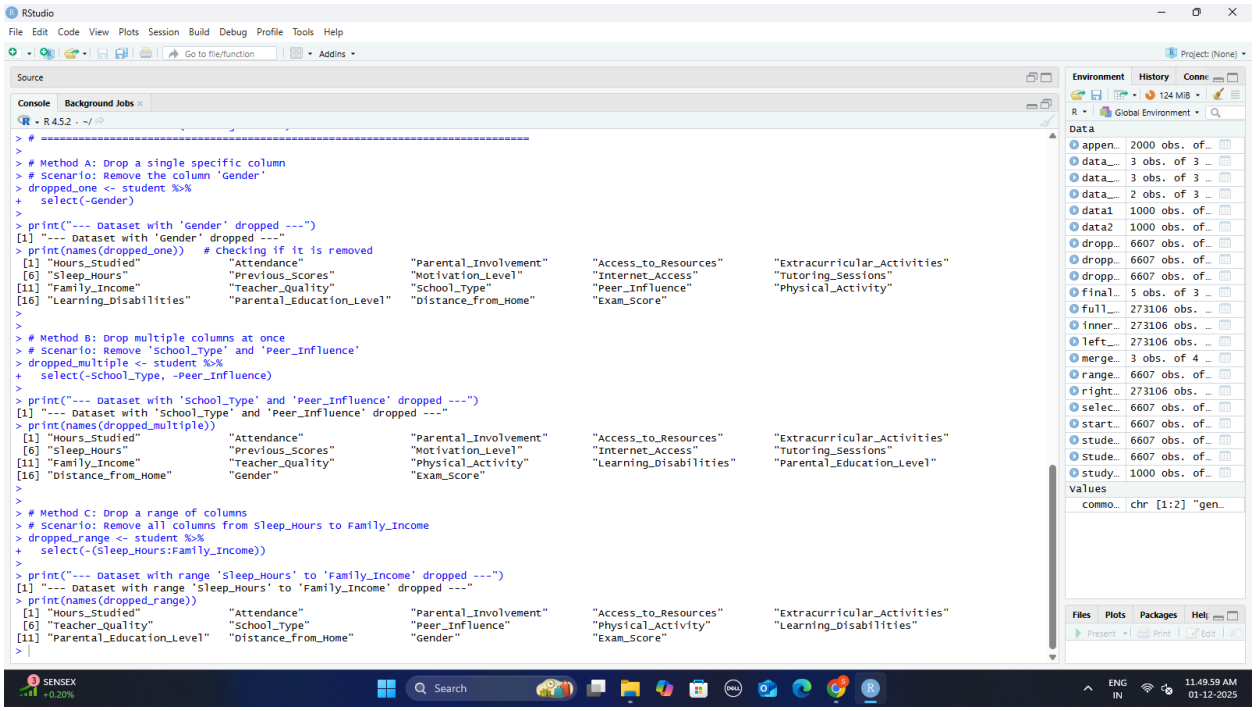

***Name :- Priya Gupta***

***Roll no. :- S081***

# Sheth L.U.J. & Sir M.V. College



```
1            23         84          67
2            19         64          61
3            24         98          74
>
>
> # Method B: Select a continuous range of adjacent columns
> # Scenario: Select from the first column to the 5th column
> # Example: Selecting columns from Hours_Studied to Extracurricular_Activities
> range_cols <- student %>%
+   select(Hours_Studied:Extracurricular_Activities)
>
> print("--- Selected Range of Columns ---")
[1] "--- Selected Range of Columns ---"
> print(head(range_cols, 3))
  Hours_Studied Attendance Parental_Involvement Access_to_Resources Extracurricular_Activities
1            23         84                  Low                High                        No
2            19         64                  Low              Medium                        No
3            24         98               Medium              Medium                       Yes
>
>
> # Method C: Select using helper functions (starts_with)
> # Scenario: Select all columns that start with the letter "P"
> # Example: Parental_Involvement, Previous_Scores, Physical_Activity, etc.
> starts_with_p <- student %>%
+   select(starts_with("P"))
>
> print("--- Selected columns starting with 'P' ---")
[1] "--- Selected columns starting with 'P' ---"
> print(head(starts_with_p, 3))
  Parental_Involvement Previous_Scores Peer_Influence Physical_Activity Parental_Education_Level
1                  Low              73       Positive                 3              High School
2                  Low              59       Negative                 4                  College
3               Medium              91        Neutral                 4              Postgraduate
>
>
> # ================================================================================
> # 3. DROPPING VARIABLES (Removing Columns)
> # ================================================================================
>
> # Method A: Drop a single specific column
> # Scenario: Remove the column 'Gender'
> dropped_one <- student %>%
```



```
> # ================================================================================
>
> # Method A: Drop a single specific column
> # Scenario: Remove the column 'Gender'
> dropped_one <- student %>%
+   select(-Gender)
>
> print("--- Dataset with 'Gender' dropped ---")
[1] "--- Dataset with 'Gender' dropped ---"
> print(names(dropped_one))   # Checking if it is removed
 [1] "Hours_Studied"              "Attendance"                 "Parental_Involvement"       "Access_to_Resources"        "Extracurricular_Activities"
 [6] "Sleep_Hours"                "Previous_Scores"            "Motivation_Level"           "Internet_Access"            "Tutoring_Sessions"
[11] "Family_Income"              "Teacher_Quality"            "School_Type"                "Peer_Influence"             "Physical_Activity"
[16] "Learning_Disabilities"      "Parental_Education_Level"   "Distance_from_Home"         "Exam_Score"
>
>
> # Method B: Drop multiple columns at once
> # Scenario: Remove 'School_Type' and 'Peer_Influence'
> dropped_multiple <- student %>%
+   select(-School_Type, -Peer_Influence)
>
> print("--- Dataset with 'School_Type' and 'Peer_Influence' dropped ---")
[1] "--- Dataset with 'School_Type' and 'Peer_Influence' dropped ---"
> print(names(dropped_multiple))
 [1] "Hours_Studied"              "Attendance"                 "Parental_Involvement"       "Access_to_Resources"        "Extracurricular_Activities"
 [6] "Sleep_Hours"                "Previous_Scores"            "Motivation_Level"           "Internet_Access"            "Tutoring_Sessions"
[11] "Family_Income"              "Teacher_Quality"            "Physical_Activity"          "Learning_Disabilities"      "Parental_Education_Level"
[16] "Distance_from_Home"         "Gender"                     "Exam_Score"
>
>
> # Method C: Drop a range of columns
> # Scenario: Remove all columns from Sleep_Hours to Family_Income
> dropped_range <- student %>%
+   select(-(Sleep_Hours:Family_Income))
>
> print("--- Dataset with range 'Sleep_Hours' to 'Family_Income' dropped ---")
[1] "--- Dataset with range 'Sleep_Hours' to 'Family_Income' dropped ---"
> print(names(dropped_range))
 [1] "Hours_Studied"              "Attendance"                 "Parental_Involvement"       "Access_to_Resources"        "Extracurricular_Activities"
 [6] "Teacher_Quality"            "School_Type"                "Peer_Influence"             "Physical_Activity"          "Learning_Disabilities"
[11] "Parental_Education_Level"   "Distance_from_Home"         "Gender"                     "Exam_Score"
>
```

**Name :- Priya Gupta**

**Roll no. :- S081**