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

Practical 12 report :- Correlation Analysis of Student Exam Scores

1. Aim

To analyze the relationship between students' study habits, attendance, previous academic performance, and current exam scores using correlation analysis and visualization techniques.

2. Dataset Description

The dataset `student_exam_scores.csv` consists of the following variables:

Column Name	Description
student_id	Unique ID for each student
hours_studied	Number of hours studied per day
sleep_hours	Average sleep hours per day
attendance_percent	Percentage of classes attended
previous_scores	Scores obtained in previous exams
exam_score	Score obtained in the current exam

S081 Priya Gupta

Name :- Priya Gupta

Roll no. :- S081

3. Results

3.1 Correlation Matrix

```
> print(cor_matrix)
                hours_studied   sleep_hours attendance_percent previous_scores exam_score
hours_studied      1.00000000  0.0778643014      -0.0313110929      0.06907820  0.7767514
sleep_hours        0.07786430  1.0000000000       0.0005719021     -0.19425104  0.1882220
attendance_percent -0.03131109  0.0005719021       1.0000000000      0.05195806  0.2257126
previous_scores    0.06907820 -0.1942510412       0.0519580567      1.00000000  0.4311047
exam_score         0.77675143  0.1882219847       0.2257126046      0.43110471  1.0000000
>
```

3.2 Observations

1. **Hours Studied:**
   - Strong positive correlation with `exam_score` (r = 0.777), showing that more study hours generally lead to better scores.
2. **Previous Scores:**
   - Moderate positive correlation with `exam_score` (r = 0.431), indicating past performance influences current results.
3. **Sleep Hours & Attendance Percent:**
   - Weak positive correlations with `exam_score` (r = 0.188 and r = 0.226), suggesting minor impact on exam performance.
4. **Other Relationships:**
   - Most other pairs have very low correlations (<0.2), showing minimal linear relationships.

S081 Priya Gupta

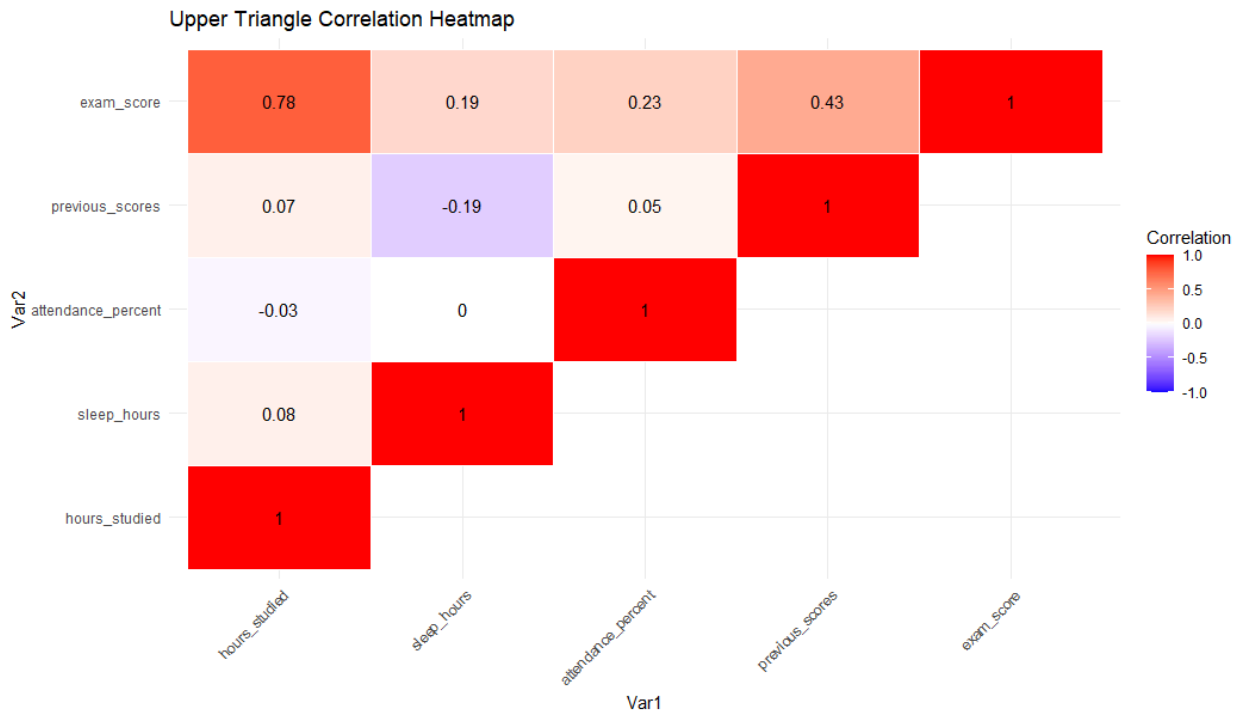
Name :- Priya Gupta

Roll no. :- S081

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

3.3 Graphical Representation

Upper-Triangle Heatmap



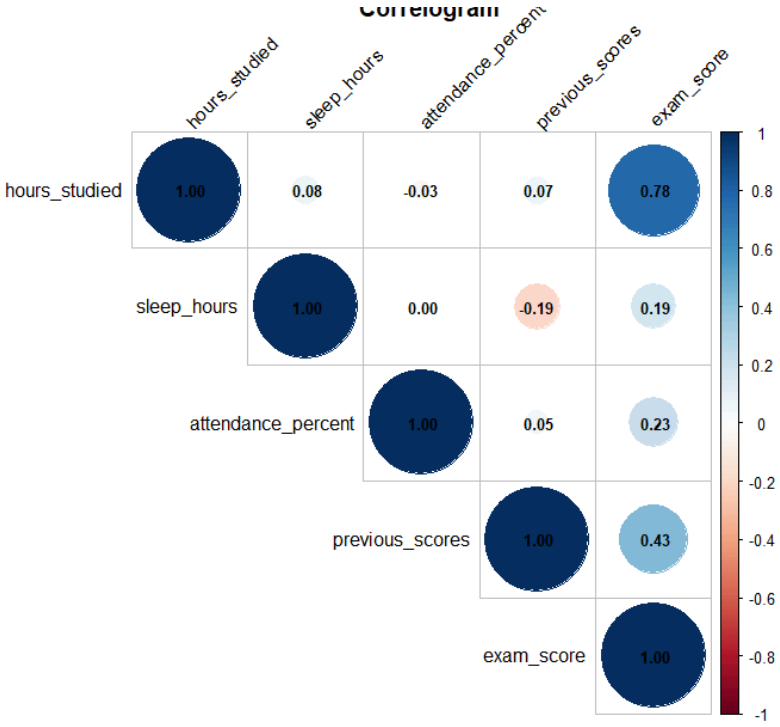
Description: Highlights correlations using color tiles. Red indicates strong positive correlation (e.g., hours studied vs exam score), white indicates weak or no correlation.

S081 PRIYA

Name :- Priya Gupta

Roll no. :- So81

Correlogram



Description: Circles represent correlations. Larger, darker circles indicate stronger relationships, making it easy to identify key influencing factors.

4. Conclusion

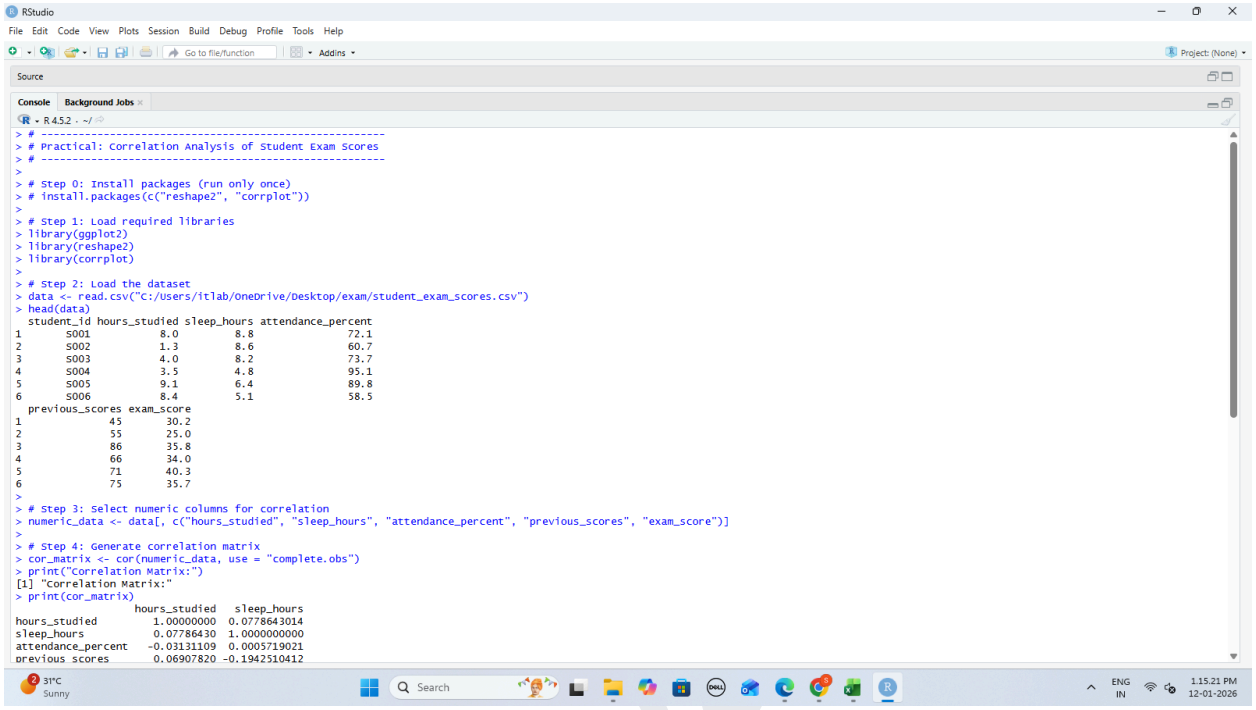
- **Hours studied per day** is the most significant factor affecting exam performance, followed by previous scores.
- Sleep hours and attendance show minimal direct impact.
- Visualizations support the correlation matrix, offering a clear understanding of factors influencing student scores.

Name :- Priya Gupta

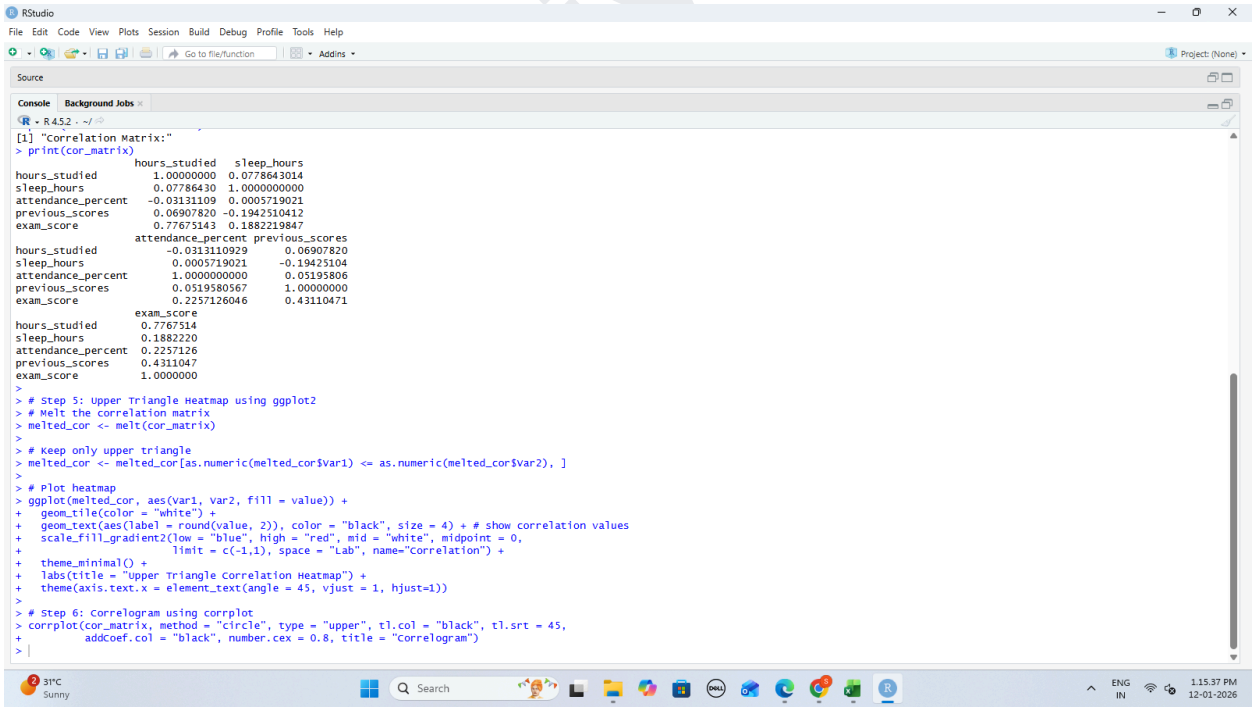
Roll no. :- So81

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

Screenshot



```
> # ------------------------------------------------------------
> # Practical: Correlation Analysis of Student Exam Scores
> # ------------------------------------------------------------
>
> # Step 0: Install packages (run only once)
> # install.packages(c("reshape2", "corrplot"))
>
> # Step 1: Load required libraries
> library(ggplot2)
> library(reshape2)
> library(corrplot)
>
> # Step 2: Load the dataset
> data <- read.csv("C:/Users/itlab/OneDrive/Desktop/exam/student_exam_scores.csv")
> head(data)
  student_id hours_studied sleep_hours attendance_percent
1       S001           8.0         8.8                72.1
2       S002           1.3         8.6                60.7
3       S003           4.0         8.2                73.7
4       S004           3.5         4.8                95.1
5       S005           9.1         6.4                89.8
6       S006           8.4         5.1                58.5
  previous_scores exam_score
1              45       30.2
2              55       25.0
3              86       35.8
4              66       34.0
5              71       40.3
6              75       35.7
>
> # Step 3: Select numeric columns for correlation
> numeric_data <- data[, c("hours_studied", "sleep_hours", "attendance_percent", "previous_scores", "exam_score")]
>
> # Step 4: Generate correlation matrix
> cor_matrix <- cor(numeric_data, use = "complete.obs")
> print("Correlation Matrix:")
[1] "Correlation Matrix:"
> print(cor_matrix)
                  hours_studied   sleep_hours
hours_studied         1.00000000  0.0778643014
sleep_hours           0.07786430  1.0000000000
attendance_percent   -0.03131109  0.0005719021
previous_scores       0.06907820 -0.1942510412
```



```
[1] "Correlation Matrix:"
> print(cor_matrix)
                  hours_studied   sleep_hours
hours_studied         1.00000000  0.0778643014
sleep_hours           0.07786430  1.0000000000
attendance_percent   -0.03131109  0.0005719021
previous_scores       0.06907820 -0.1942510412
exam_score            0.77675143  0.1882219847
                  attendance_percent previous_scores
hours_studied          -0.0313110929      0.06907820
sleep_hours             0.0005719021     -0.19425104
attendance_percent      1.0000000000      0.05195806
previous_scores         0.0519580567      1.00000000
exam_score              0.2257126046      0.43110471
                  exam_score
hours_studied       0.7767514
sleep_hours         0.1882220
attendance_percent  0.2257126
previous_scores     0.4311047
exam_score          1.0000000
>
> # Step 5: Upper Triangle Heatmap using ggplot2
> # Melt the correlation matrix
> melted_cor <- melt(cor_matrix)
>
> # Keep only upper triangle
> melted_cor <- melted_cor[as.numeric(melted_cor$Var1) <= as.numeric(melted_cor$Var2), ]
>
> # Plot heatmap
> ggplot(melted_cor, aes(Var1, Var2, fill = value)) +
+   geom_tile(color = "white") +
+   geom_text(aes(label = round(value, 2)), color = "black", size = 4) + # show correlation values
+   scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0,
+                        limit = c(-1,1), space = "Lab", name="Correlation") +
+   theme_minimal() +
+   labs(title = "Upper Triangle Correlation Heatmap") +
+   theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
>
> # Step 6: Correlogram using corrplot
> corrplot(cor_matrix, method = "circle", type = "upper", tl.col = "black", tl.srt = 45,
+          addCoef.col = "black", number.cex = 0.8, title = "Correlogram")
>
```

Name :- Priya Gupta

Roll no. :- So81