## Project Title:

Student Performance Analysis Using SQL

## Project Overview:

This project analyzes the academic performance of 200 students using SQL. The goal is to understand how study hours, sleep, and attendance impact exam scores, identify top and bottom performers, and uncover actionable insights for improving student outcomes.

## Dataset:

- Source: Simulated student performance dataset
- Number of Records: 200 students
- Columns:
- student_id – Unique student identifier
- hours_studied – Daily study hours
- sleep_hours – Daily sleep hours
- attendance_percent – Class attendance (%)
- previous_scores – Previous assessment scores
- exam_score – Current exam score

## Objectives:

1. Identify top and bottom performers.
2. Understand patterns in study habits, sleep, and attendance.
3. Generate actionable insights for students and educators.
4. Demonstrate SQL-based data aggregation and analysis.

## Tools & Skills Used:

- SQL: MySQL Workbench, data aggregation, grouping, filtering, and ranking.
- Data Analysis: Exploring distributions, calculating averages, ranking top/bottom performers, analyzing trends.
- Insights Reporting: Translating SQL results into actionable findings.

## Step-by-Step Approach:

### 1. Data Loading & Table Creation

- Imported CSV dataset into MySQL Workbench using `LOAD DATA INFILE`.
- Created a structured table `student_scores` for analysis:

```
CREATE TABLE student_scores (
    student_id VARCHAR(10) PRIMARY KEY,
    hours_studied DECIMAL(4,1),
    sleep_hours DECIMAL(4,1),
    attendance_percent DECIMAL(5,1),
```

```
    previous_scores INT,
    exam_score DECIMAL(4,1)
);
```

## 2. Exploratory Analysis Using SQL

• Top & Bottom Performers:

```sql
-- Top 5 students
SELECT student_id, exam_score
FROM student_scores
ORDER BY exam_score DESC
LIMIT 5;

-- Bottom 5 students
SELECT student_id, exam_score
FROM student_scores
ORDER BY exam_score ASC
LIMIT 5;
```

• Sleep Hours vs Exam Score:

```sql
SELECT ROUND(sleep_hours) AS sleep_group,
       AVG(exam_score) AS avg_exam_score,
       COUNT(*) AS student_count
FROM student_scores
GROUP BY sleep_group
ORDER BY sleep_group DESC;
```

• Attendance vs Exam Score:

```sql
SELECT
    CASE
        WHEN attendance_percent >= 90 THEN '90+'
        WHEN attendance_percent >= 75 THEN '75-89'
        ELSE 'Below 75'
    END AS attendance_group,
    AVG(exam_score) AS avg_exam_score,
    COUNT(*) AS student_count
FROM student_scores
GROUP BY attendance_group
ORDER BY avg_exam_score DESC;
```

• Study Hours vs Exam Score:

```sql
SELECT hours_studied,
       AVG(exam_score) AS avg_exam_score,
       COUNT(*) AS student_count
```

```
FROM student_scores
GROUP BY hours_studied
ORDER BY hours_studied DESC;
```

### 3. Key Findings & Insights

1. Top Exam Scores: The highest exam score is 51.3%, showing generally low performance across the cohort.
2. Bottom Exam Scores: Students scoring below 20% had very low study hours and attendance.
3. Study Hours Impact: Students studying 10–12 hours performed significantly better than those studying fewer hours.
4. Sleep Hours Impact: 7–8 hours of sleep correlated with slightly higher scores; too little or too much sleep had lower averages.
5. Attendance Impact: High attendance (>90%) generally correlated with better performance but was less influential than study hours.

### 4. Actionable Recommendations

- Encourage students to increase study hours, particularly those below 5h/day.
- Promote consistent attendance to maintain engagement.
- Monitor students with low study hours or attendance for early intervention.
- Track trends using SQL queries to identify at-risk students continuously.

### Conclusion

This SQL-driven analysis demonstrates how to extract insights, rank students, and identify key performance drivers using simple SQL queries. Study hours are the strongest predictor of success, while sleep and attendance play supportive roles.

### Skills Demonstrated

- SQL table creation and data import
- Data aggregation, grouping, and filtering
- Ranking and identifying top/bottom performers
- Generating actionable insights from raw data