Factors Influencing Student Exam Performance: An Analytical Study

Executive Summary

This project explores which factors most strongly influence student exam scores by analyzing a comprehensive simulated dataset from the table `student_performance`. The analysis uses SQL queries to examine how hours studied, participation in extracurricular activities, and other aspects such as attendance, sleep hours, tutoring, and teacher quality affect outcomes. The results support actionable recommendations for students, teachers, and policymakers seeking to boost academic achievement.

Project Objectives

- Quantify the relationship between **hours studied**, extracurricular participation, and exam scores.
- Analyze the impact of supporting factors such as attendance, sleep, and tutoring on student performance.
- Surface evidence-based actions to improve learning outcomes.

Technologies & Tools Used

- **Database**: PostgreSQL
- **Analysis Platform**: Jupyter Notebook (`notebook.ipynb`)
- **SQL Techniques**: Aggregation, grouping, filtering, ranking

Dataset Description

Table: `student_performance`

Column	Description	Type
`attendance`	% of classes attended	float

`extracurricular_activities`	'Yes'/'No' participation	varchar
`sleep_hours`	Avg. hours of sleep/night	float
`tutoring_sessions`	Monthly tutoring sessions attended	integer
`teacher_quality`	Teacher quality rating (Low/Medium/High)	varchar
`exam_score`	Final exam score	float
`hours_studied`	Number of study hours per week	integer

Approach & Methodology

1. **Data Exploration:**
    - Inspected available metrics; focused on identifying high-impact factors.
2. **SQL Aggregation:**
    - Averaged exam scores by hours studied and extracurricular participation for students dedicating more than 10 hours/week to study and who were engaged in extracurriculars:

```
SELECT
  hours_studied,
  ROUND(AVG(exam_score)::numeric, 2) AS avg_exam_score
FROM student_performance
WHERE hours_studied > 10
  AND extracurricular_activities = 'Yes'
GROUP BY hours_studied
ORDER BY hours_studied DESC;
```

3. **Pattern Analysis:**
    - Compared score distributions across a range of study hours and extracurricular involvement.
    - Considered additional context such as attendance, tutoring, and sleep for top scorers.

Results & Insights

Exam Score by Hours Studied (With Extracurricular Participation)

Hours Studied	Avg Exam Score
43	78.00
39	75.00
38	73.50
37	73.00
...	...
11	65.15

- **Key trend:** There is a clear positive association between higher hours studied (>10 per week) and average exam scores for students involved in extracurriculars.
- The highest average scores (78.00) were for students studying 43 hours/week.

Additional Analyses

- **Grouped by study-hour brackets:**
Hours Studied Range	Avg Exam Score
16+ hours	67.92
11–15 hours	65.20
6–10 hours	64.23
1–5 hours	62.63

- **Top Performing Students:**
  - Most top-10 scorers had high attendance (≥90%), studied >18 hours/week, slept 6–9 hours/night, and usually had several tutoring sessions monthly.

Challenges & Solutions

Challenge	Solution
Text/numeric type variance	Ensured proper casting/rounding
Missing/irregular records	Focused on complete, relevant fields

Conclusion

Hours studied is the single most important factor associated with high exam performance, especially for students who also participate in extracurricular activities. Other contributors to high scores include strong attendance, regular tutoring, and adequate sleep. These results are directly actionable for academic planning and student support interventions.

Next Steps

- Expand analysis to study non-linear effects and optimal study/sleep balance.
- Explore interaction effects for variables like teacher quality or tutoring.
- Visualize trends with charts/graphs for curriculum planning and student advising.

See the accompanying Jupyter notebook (notebook.ipynb) for full queries, data outputs, and reproducibility.