

In today's fast-paced and competitive educational environment, understanding the factors that influence student success is more important than ever. Just like the transport system in a bustling city like London must adapt to serve its residents, schools and educators must adapt to meet the needs of students. In this project, we will take a deep dive into a dataset containing rich details about various aspects of student life, such as hours studied, sleep patterns, attendance, and more, to uncover what truly impacts exam performance.

The dataset we'll be working with includes a wide range of factors influencing student performance. By analyzing this data, we'll be able to identify key drivers of success and provide insights that could help students, teachers, and policymakers make informed decisions. The table we'll use for this project is called `student_performance` and includes the following data:

Column	Definition	Data type
<code>attendance</code>	Percentage of classes attended	<code>float</code>
<code>extracurricular_activities</code>	Participation in extracurricular activities	<code>varchar</code> (Yes, No)
<code>sleep_hours</code>	Average number of hours of sleep per night	<code>float</code>
<code>tutoring_sessions</code>	Number of tutoring sessions attended per month	<code>integer</code>
<code>teacher_quality</code>	Quality of the teachers	<code>varchar</code> (Low, Medium, High)
<code>exam_score</code>	Final exam score	<code>float</code>

You will execute SQL queries to answer three questions, as listed in the instructions.

Projects Data    DataFrame as `avg_exam_score_by_study_and_exti`

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

index	...	↑↓	hours_studied	...	↑↓	avg_exam_score	...	↑↓
			0			43		
			1			39		
			2			38		
			3			37		
			4			36	70.4285714	
			5			35	72.3	
			6			34	71.1	
			7			33	70.3333333	
			8			32	71.	
			9			31	70.5531914	
			10			30	71.4328358	
			11			29	70.256097	
			12			28	69.8256880	
			13			27	69.7768595	
			14			26	68.801369	
			15			25		

Rows: 30



Projects Data DataFrame as avg\_exam\_score\_by\_hours\_studied.

```
-- avg_exam_score_by_hours_studied_range
SELECT
    CASE WHEN (hours_studied >= 1 AND hours_studied <= 5) THEN '1-5 hours'
    WHEN (hours_studied >= 6 AND hours_studied <= 10) THEN '6-10 hours'
    WHEN (hours_studied >= 11 AND hours_studied <= 15) THEN '11-15 hours'
    ELSE '16+ hours' END AS hours_studied_range,
    AVG(exam_score) AS avg_exam_score
FROM student_performance
GROUP BY hours_studied_range
ORDER BY avg_exam_score DESC;
```

index	...	↑↓	hours_studied_range	...	↑↓	avg_exam_score	...
			0			16+ hours	67.9233633
			1			11-15 hours	65.2043859
			2			6-10 hours	64.2254901
			3			1-5 hours	62.6271186

Rows: 4



Projects Data DataFrame as student\_exam\_ranking

```
-- student_exam_ranking
SELECT
    attendance,
    hours_studied,
    sleep_hours,
    tutoring_sessions,
    DENSE_RANK() OVER(ORDER BY exam_score DESC) AS exam_rank
FROM student_performance
LIMIT 30;
```

i.	...	↑↓	atten...	...	↑↓	hours_studi...	...	↑↓	sleep_h...	...	↑↓	tutoring_sessions	...	↑↓	exa...	...	↑↓
		0			98			27			6			5			