EDTS SQL PROJECT SEASON 4

Project Instructions: Student Performance Dataset Analysis

You are required to conduct an in-depth analysis of the Student Performance dataset provided. The goal of this exercise is to help you apply your SQL knowledge in a real-world context and to gain practical experience in data exploration, querying, and insight generation.
Please read the instructions below carefully before you begin:

Instructions for Analysis

1. Use Any Available Data Tools

Begin your analysis by exploring the dataset using any data analysis tools available to you. This may include Excel, Power BI, Python (Pandas), or any tool you are comfortable with before transitioning to SQL. The objective is to first understand the structure and quality of the data.

2. Handle Data Inconsistencies Wisely

If you encounter inconsistencies, missing values, or unclear entries that might affect your ability to query effectively in SQL, you are permitted and encouraged to clean or transform the dataset beforehand. This may involve:
o Splitting or formatting columns
o Handling missing or duplicate data
o Normalizing fields or converting data types
Ensure your final dataset is suitable for relational querying.

3. Use the Questions as Guides, Not Limits

You have been given a set of questions to guide your thinking and querying. However, you are not limited to only those questions. Use them as a starting point to explore trends, patterns, and relationships within the dataset.

4. Be Creative and Insightful

We encourage creativity and critical thinking. Go beyond surface-level observations and look for meaningful insights such as:
o Influence of study habits on exam scores
o The role of parental involvement and education
o Impact of extracurricular activities, sleep, or physical activity

o Comparisons across gender, school type, or income levels
Use charts, summaries, and well-structured queries to support your analysis.

5. Document Your Process

Keep notes of the steps you took to clean or transform the data (if any), the tools used, and your thought process when answering each question. This will help you articulate your approach during presentations or reviews.

Student Performance SQL Project Questions.

Part 1: SQL Project Questions

1. How many students are included in the dataset?
2. What is the average Exam Score across all students?
3. Find the average Exam Score grouped by Gender.
4. What is the correlation between Hours_Studied and Exam_Score?
5. Which School_Type (Public or Private) has the highest average Exam Score?
6. How does Attendance percentage affect Exam Scores (e.g., average scores by attendance ranges)?
7. Compare Exam Scores between students with and without Tutoring_Sessions.
8. Find the distribution of Exam Scores by Parental_Education_Level.
9. How does Family_Income category affect Exam Scores?
10. Compare Exam Scores of students with and without Extracurricular_Activities.
11. Does Internet_Access show any relationship with Exam Scores?
12. What is the impact of Motivation_Level on Exam Scores?
13. Identify students with Learning Disabilities and compare their average Exam Score to others.
14. Which factor (Hours_Studied, Sleep_Hours, and Physical Activity) shows the strongest link with Exam Scores?
15. List the top 10 highest scoring students and their key attributes (Hours_Studied, Attendance, Motivation_Level, etc.).

Part 2: Student Performance Dataset Data Profile

Total Columns: 20

Hours_Studied: Number – Hours the student spent studying.

Attendance: Number – Attendance percentage in class.

Parental_Involvement: Text – Level of parental involvement (Low, Medium, High).

Access_to_Resources: Text – Availability of study resources (Low, Medium, High).

Extracurricular_Activities: Text – Participation in extracurricular activities (Yes/No).

Sleep_Hours: Number – Average number of sleep hours per night.

Previous_Scores: Number – Previous exam scores of the student.

Motivation_Level: Text – Self-reported motivation level (Low, Medium, High).

Internet_Access: Text – Whether the student has internet access (Yes/No).

Tutoring_Sessions: Number – Number of tutoring sessions attended.

Family_Income: Text – Family's income level (Low, Medium, High).

Teacher_Quality: Text – Teacher's quality rating (Low, Medium, High).

School_Type: Text – Type of school (Public or Private).

Peer_Influence: Text – Influence of peers (Positive, Neutral, Negative).

Physical_Activity: Number – Hours per week of physical activity.

Learning_Disabilities: Text – Whether the student has learning disabilities (Yes/No).

Parental_Education_Level: Text – Highest education level of parents.

Distance_from_Home: Text – Distance to school (Near, Moderate, Far).

Gender: Text – Gender of the student (Male/Female).

Exam_Score: Number – Final exam score of the student.

Submission Guidelines

- ✓ Submit a .sql file showing your SQL queries and answers.
- ✓ Submit a visualization file too either in Excel or PowerBI
- ✓ Submit also a .ppt file for review.
- ✓ Include a short summary of your findings in bullet points or paragraph format.
- ✓ Submit any data preparation files if you transformed the dataset prior to SQL querying.
- ✓ Submit to my mail on or before the dateline: *legallheroes@gmail.com*

Good luck!