STUDENT PERFORMANCE ANALYSIS

GROUP – 1

- Harsha
- Irene
- Ardhra
- Parvathy
- Abhiram
- Adarsh

INTRODUCTION

This report provides an in-depth analysis of factors influencing student performance in exams, using a dataset that captures various academic, personal, and socio-economic variables. The dataset includes key attributes such as hours studied, attendance, and parental involvement, offering insights into how these aspects impact academic success.

Beyond academic habits, the dataset examines other crucial elements, such as the influence of peer relationships, motivation levels, and physical activity, all of which can play a role in shaping academic performance.

Student performance analysis is a crucial aspect of educational institutions. The findings from this report will serve as valuable insights for educators, parents, and policymakers seeking to improve academic achievement through targeted strategies that address both in-school and external factors.

AIM

To develop a comprehensive MySQL database system for effectively storing, analyzing, and visualizing student performance data, enabling educators to identify trends, assess teaching effectiveness, and provide targeted support to enhance overall student outcomes.

OBJECTIVE

1. Data Collection:
   Gather relevant student data, including demographic information, academic records, attendance data, and behavioral metrics.

Utilize various sources such as school management systems, gradebooks, attendance records, and surveys.

2. Data Cleaning:
   Identify and correct errors, inconsistencies, or missing values within the collected data. Standardize data formats and ensure data integrity. Transform raw data into a suitable format for analysis.

3. SQL Queries
   Designed and executed SQL queries to extract relevant information from the database. This involved a range of SQL operations, including SELECT, GROUP BY, and aggregation functions.

DATA ANALYSIS:

1. **Understanding the Data:**
   Before diving into the analysis, ensure you have a well-structured MySQL database containing relevant student performance data. This might include:

- **Student Information:** gender, grade level, previous grade level
- **Performance Data:** Grades for assignments, quizzes, tests, final exams, attendance
- **Other Involvement :** Parental Involvement, peer influence, teacher quality.

2. **Grade Performance:**

Compare the previous obtained scores with the now obtained scores

.

3. **Performance by Gender:**
   Compare average grades, motivation levels, attendance, parental involvement, and other statistical factors for male and female students.

4. **Relationship between two constraints :**
   Calculate the correlation between grades and attendance, grades and parental involvement, attendance and motivation level and other constraints.

DATA OVERVIEW

Column name	Datatype
Attendance	int
Parental_Involvement	varchar
Access_to_Resources	varchar
Extracurricular_Activities	Varchar
Sleep_Hours	Int
Previous_Scores	Int
Motivation_Level	Varchar
Internet_Access	Varchar
Tutoring_Sessions	Int
Family_Income	Varchar
Teacher_Quality	Varchar
School_Type	Varchar
Peer_Influence	Varchar

Physical_Activity	Int
Learning_Disabilities	Varchar
Parental_Education_Level	Varchar
Distance_from_Home	Varchar
Gender	Varchar
Exam_Score	int
Hours_studied	int

Attribute Description

- Hours_Studied - Number of hours spent studying per week.
- Attendance - Percentage of classes attended.
- Parental_Involvement - Level of parental involvement in the student's education (Low, Medium, High).
- Access_to_Resources - Availability of educational resources (Low, Medium, High).
- Extracurricular_Activities - Participation in extracurricular activities (Yes, No).
- Sleep_Hours - Average number of hours of sleep per night.
- Previous_Scores - Scores from previous exams.
- Motivation_Level - Student's level of motivation (Low, Medium, High).
- Internet_Access - Availability of internet access (Yes, No).
- Tutoring_Sessions - Number of tutoring sessions attended per month.
- Family_Income - Family income level (Low, Medium, High).
- Teacher_Quality - Quality of the teachers (Low, Medium, High).
- School_Type - Type of school attended (Public, Private).
- Peer_Influence - Influence of peers on academic performance (Positive, Neutral, Negative).

- Physical_Activity - Average number of hours of physical activity per week.
- Learning_Disabilities - Presence of learning disabilities (Yes, No).
- Parental_Education_Level - Highest education level of parents (High School, College, Postgraduate).
- Distance_from_Home - Distance from home to school (Near, Moderate, Far).
- Gender - Gender of the student (Male, Female).
- Exam_Score - Final exam score.

QUERIES

1. To find the total number of students
   SELECT COUNT(*) FROM Student_factors_project;

2. To find number of males and females
   SELECT COUNT(*)FROM Student_factors_project GROUP BY Gender;

3. What is the average number of hours studied by students
   q- SELECT AVG(Hours_Studied) FROM Student_factors_project

4. What is the average number of hours studied by students, categorized by gender?
   q- SELECT Gender, AVG(Hours_Studied) FROM Student_factors_project GROUP BY Gender;

5. Students with access to resources

q-SELECT Access_To_Resources ,count(*) FROM student_factors_project GROUP BY Access_To_Resources;

6. What is the average number of hours of sleep for students who have a high level of motivation?
   q-SELECT AVG(Sleep_Hours) FROM Student_factors_project WHERE Motivation_level = 'High';

7. What is the total sum of exam scores for students, grouped by their school type?
   q-SELECT School_Type, SUM(Exam_Score)
   FROM Student_factors_project
   GROUP BY School_Type;

8. What is the highest exam score achieved by any student in the dataset?
   q-SELECT MAX(Exam_Score)
   FROM Student_factors_project;

9. What is the max of exam scores for students, grouped by their school type?
   q-SELECT School_Type, max(Exam_Score)
   FROM Student_factors_project
   GROUP BY School_Type;

10. What is the average exam score of students who have a high number of tutoring sessions?
    q- SELECT AVG(Exam_Score) FROM Student_factors_project WHERE Tutoring_Sessions = 'High';

11. -What are the average exam scores for students based on their school type, and how do these scores rank from highest to lowest?
    q- SELECT School_Type, AVG(Exam_Score)
    FROM Student_factors_project
    GROUP BY School_Type
    ORDER BY AVG(Exam_Score) DESC;

12. How many students without learning disabilities have a low level of motivation?
    q-select count(*) Exam_Scores from Student_factors_project where Learning_Disabilities='No' AND Motivation_Level='Low';

13. How many students have high parental involvement but low access to educational resources?
    q- SELECT COUNT(*) FROM Student_factors_project WHERE Parental_Involvement = 'High' AND Access_to_Resources = 'Low';

14. How many female students are there in the dataset with a specified parental involvement?
    q-select count(*) Parental_Involvement from Student_factors_project where Gender='female';

15. How many male students are there in the dataset with a specified parental involvement?

q-select count(*) Parental_Involvement from Student_factors_project where Gender='male';

16. How many students have low parental involvement but high access to educational resources?
q-SELECT COUNT(*) FROM Student_factors_project WHERE Parental_Involvement = 'Low' AND Access_to_Resources = 'High';

17. How many students have low internet access and high participation in extracurricular activities?
q-SELECT COUNT(*)
FROM Student_factors_project
WHERE Internet_Access = 'Low' AND Extracurricular_Activities = 'High';

18. What is the average amount of physical activity reported by students, categorized by gender?
q-select Gender,avg(Physical_Activity) from student_factors_project group by Gender;

19. What is the average exam score for students attending different types of schools (public and private)?
q- SELECT School_Type, AVG(Exam_Score) FROM Student_factors_project GROUP BY School_Type;

20. What is the average exam score for students based on their family income levels?
    q- SELECT Family_Income, AVG(Exam_Score) FROM Student_factors_project GROUP BY Family_Income;

21. How many students have no internet access and low parental involvement?
    q-select count(*) Exam_Scores from Student_factors_project where Internet_Access='No' AND Parental_Involvement='low';

22. What is the average exam score of students who participate in extracurricular activities?
    q-select AVG (Exam_Score) from Student_factors_project where Extracurricular_Activities='yes';

23. 23.What is the average number of hours studied by students, grouped by different levels of parental involvement
    q-SELECT AVG(Hours_Studied),Parental_Involvement
    FROM Student_factors_project
    GROUP BY Parental_Involvement;

24. What is the average exam score of students, grouped by different levels of motivation
    q-SELECT AVG(Exam_Score),Motivation_Level
    FROM Student_factors_project
    GROUP BY Motivation_Level;

25. What is the difference and percentage change between a student's final exam score and their previous exam scores.

```
q-SELECT Previous_Scores, Exam_Score,
  (Exam_Score - Previous_Scores) AS Score_Difference,
 (Exam_Score / Previous_Scores) * 100 AS Percentage_Change
  FROM Student_factors_project;
```

26. Classifying students on basis of motivation level

```
q-SELECT Motivation_Level, AVG(Exam_Score) AS
Average_Exam_Score,
 AVG(Hours_Studied) AS Avg_Hours_Studied,
 AVG(Attendance) AS Avg_Attendance,
 AVG(Sleep_Hours) AS Avg_Sleep_Hours,
 AVG(Previous_Scores) AS Avg_Previous_Scores
FROM Student_factors_project
GROUP BY Motivation_Level;
```

CONCLUSION

1. **Total number of students**
   There are a total 6607 students in the dataset

2. **Total number of females and males**
   Females: 2793
   Males: 3814

3. **Average amount of hours studied by students**
   An approx. of 20 hours (19.9753) is the average amount of time spend by students for studying .

4. **Access to resources by the students**
   High : 1975
   Medium : 3319
   Low : 1313

5. **Students involved in extracurricular activities**
   Students doing extracurricular activities 3938, not involved in an extracurricular activities 2669.

6. **Students with parent involvement when access to resources is high**
   Number of students with high parent involvement is 568, parent involvement medium is 993 and parent involvement low for 414 students.

7. **Scores on basis of school type**
   The average exam score based on the school type is
   Public : 67.2129
   Private : 67.2877

8. **Amount of sleep**

The average amount of sleep by the students is 7 hours .

9. **Students classification on basis of motivation level**

Motivation_Level	Average_Exam_Score	Avg_Hours_Studied	Avg_Attendance	Avg_Sleep_Hours	Avg_Previous_Scores
Low	66.7522	19.9561	79.9344	7.0263	75.2003
Medium	67.3306	20.0830	80.1065	7.0301	74.7929
High	67.7043	19.7301	79.7127	7.0303	75.5853

10. **Parental involvement in different school types**

Low Parental Involvement results in similar performance between public and private schools, with only a marginal difference in exam scores.

Medium and High Parental Involvement levels lead to progressively better exam performance, with private schools showing a slight, but consistent, advantage over public schools at all involvement levels.

Public Schools appear to close the gap with private schools at High Parental Involvement levels, showing that strong parental support can mitigate the differences between school types.

In summary, the dataset contains a total of 6,607 students, with a gender distribution of 2,793 females and 3,814 males. Students study an average of approximately 20 hours per week (19.98 hours), and their access to resources is categorized as high (1,975), medium (3,319), and low (1,313). Regarding extracurricular activities, 3,938 students are involved, while 2,669 are not. Among students with high resource access, parental involvement is reported as high for 568 students, medium for 993, and low for 414. The average exam scores reveal a slight advantage for private schools (67.29) over public schools (67.21). Students average about 7 hours of sleep. Additionally, low

parental involvement results in similar performance between public and private schools, while medium and high involvement levels enhance performance, with private schools consistently performing better. Public schools show potential to close the gap with private schools at high parental involvement levels, underscoring the importance of parental support in academic success.

To improve student performance, focus on increasing resource access for students with low availability and fostering greater parental involvement, especially in public schools. Additionally, promoting extracurricular activities and ensuring a balance between study, sleep, and involvement can enhance overall academic success.