**📊 Student Performance Analysis Project Report**

**📁 Project Title:**

**“Analyzing Factors Affecting Student Exam Performance”**

**🧠 Objective:**

The aim of this project is to explore and analyze the impact of various socio-academic and personal factors on students’ exam scores using a real-world dataset. This analysis helps uncover which parameters significantly influence academic performance and can guide stakeholders (schools, parents, policymakers) to make data-driven decisions.

**📚 Dataset Overview:**

The dataset comprises **6607 student records** with **20 attributes**, including features such as:

* **Academic:** Hours\_Studied, Attendance, Previous\_Scores, Exam\_Score
* **Lifestyle:** Sleep\_Hours, Physical\_Activity
* **Socioeconomic & Family:** Family\_Income, Parental\_Education\_Level, Parental\_Involvement
* **Resources & Support:** Access\_to\_Resources, Tutoring\_Sessions, Teacher\_Quality
* **Demographic:** Gender, School\_Type, Distance\_from\_Home

**🧼 Data Preprocessing:**

* ✅ Removed duplicate records (final shape: **6607 rows**, **20 columns**).
* ❗ Missing values found in:
  + Teacher\_Quality (78 missing)
  + Parental\_Education\_Level (90 missing)
  + Distance\_from\_Home (67 missing)
* Suggested treatment:
  + Impute missing categorical data with mode or use predictive imputation.
* Data types: Mix of **numerical** and **categorical** features.

**📊 Descriptive Statistics:**

* **Average Hours Studied:** ~20 hrs/week
* **Average Attendance:** ~80%
* **Average Exam Score:** ~67.2
* **Max Exam Score:** 101
* **Students scoring above 85:** 34

**📈 Feature Correlation with Exam Score:**

| **Feature** | **Correlation** |
| --- | --- |
| Attendance | **+0.58** |
| Hours\_Studied | +0.45 |
| Previous\_Scores | +0.18 |
| Tutoring\_Sessions | +0.16 |
| Physical\_Activity | +0.03 |
| Sleep\_Hours | -0.02 |

📌 **Key Insight:**  
**Attendance** and **Hours\_Studied** show the **strongest positive correlation** with Exam\_Score.

**👨‍👩‍👧 Gender-Based Performance:**

| **Gender** | **Avg Exam Score** |
| --- | --- |
| Female | 67.24 |
| Male | 67.23 |

✅ Performance across genders is nearly identical.

**🔍 Insights:**

1. 📈 **Attendance is the top predictor** of performance.
2. ⏰ **Hours studied per week** has a strong effect on scores.
3. 🧠 Prior academic performance (Previous\_Scores) gives a fair indication of future success.
4. 📉 **Sleep\_Hours** shows negligible (and slightly negative) correlation—may reflect trade-off between study hours and sleep.
5. 📚 **Tutoring\_Sessions** have a small but positive impact.
6. ⚖️ **Extracurricular\_Activities, Internet\_Access, and School\_Type** require further categorical analysis (e.g., with boxplots or encoding for regression).

**🧪 Next Steps (Optional Enhancements):**

* **Data Cleaning:** Impute missing values or drop rows based on use-case.
* **Encoding:** Apply Label Encoding or One-Hot Encoding to categorical variables.
* **Feature Engineering:** Combine/derive features (e.g., “Study-to-Sleep ratio”).
* **Modeling:** Use ML models like Linear Regression, Decision Trees, or Random Forest to predict Exam Scores.
* **Visualization:** Create bar charts, heatmaps, or boxplots for deeper categorical insights.

**🖥️ Code Used:**

Let me know if you’d like the **full combined code** including:

* Data loading
* Cleaning
* Descriptive analysis
* Visualizations
* Model (optional)

**📌 Conclusion:**

This project highlights that **consistency in attendance and time devoted to studying are key indicators** of student success. While external support like tutoring and resources do help, intrinsic factors like motivation and habits play a major role in shaping outcomes. This analysis can help design interventions for struggling students or identify areas of improvement in education delivery.

OUTPUT

First 5 rows:

Hours\_Studied Attendance Parental\_Involvement Access\_to\_Resources \

0 23 84 Low High

1 19 64 Low Medium

2 24 98 Medium Medium

3 29 89 Low Medium

4 19 92 Medium Medium

Extracurricular\_Activities Sleep\_Hours Previous\_Scores Motivation\_Level \

0 No 7 73 Low

1 No 8 59 Low

2 Yes 7 91 Medium

3 Yes 8 98 Medium

4 Yes 6 65 Medium

Internet\_Access Tutoring\_Sessions Family\_Income Teacher\_Quality \

0 Yes 0 Low Medium

1 Yes 2 Medium Medium

2 Yes 2 Medium Medium

3 Yes 1 Medium Medium

4 Yes 3 Medium High

School\_Type Peer\_Influence Physical\_Activity Learning\_Disabilities \

0 Public Positive 3 No

1 Public Negative 4 No

2 Public Neutral 4 No

3 Public Negative 4 No

4 Public Neutral 4 No

Parental\_Education\_Level Distance\_from\_Home Gender Exam\_Score

0 High School Near Male 67

1 College Moderate Female 61

2 Postgraduate Near Male 74

3 High School Moderate Male 71

4 College Near Female 70

Dataset Info:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 6607 entries, 0 to 6606

Data columns (total 20 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Hours\_Studied 6607 non-null int64

1 Attendance 6607 non-null int64

2 Parental\_Involvement 6607 non-null object

3 Access\_to\_Resources 6607 non-null object

4 Extracurricular\_Activities 6607 non-null object

5 Sleep\_Hours 6607 non-null int64

6 Previous\_Scores 6607 non-null int64

7 Motivation\_Level 6607 non-null object

8 Internet\_Access 6607 non-null object

9 Tutoring\_Sessions 6607 non-null int64

10 Family\_Income 6607 non-null object

11 Teacher\_Quality 6529 non-null object

12 School\_Type 6607 non-null object

13 Peer\_Influence 6607 non-null object

14 Physical\_Activity 6607 non-null int64

15 Learning\_Disabilities 6607 non-null object

16 Parental\_Education\_Level 6517 non-null object

17 Distance\_from\_Home 6540 non-null object

18 Gender 6607 non-null object

19 Exam\_Score 6607 non-null int64

dtypes: int64(7), object(13)

memory usage: 1.0+ MB

None

Missing values per column:

Hours\_Studied 0

Attendance 0

Parental\_Involvement 0

Access\_to\_Resources 0

Extracurricular\_Activities 0

Sleep\_Hours 0

Previous\_Scores 0

Motivation\_Level 0

Internet\_Access 0

Tutoring\_Sessions 0

Family\_Income 0

Teacher\_Quality 78

School\_Type 0

Peer\_Influence 0

Physical\_Activity 0

Learning\_Disabilities 0

Parental\_Education\_Level 90

Distance\_from\_Home 67

Gender 0

Exam\_Score 0

dtype: int64

Shape after removing duplicates: (6607, 20)

Summary Statistics:

Hours\_Studied Attendance Sleep\_Hours Previous\_Scores \

count 6607.000000 6607.000000 6607.00000 6607.000000

mean 19.975329 79.977448 7.02906 75.070531

std 5.990594 11.547475 1.46812 14.399784

min 1.000000 60.000000 4.00000 50.000000

25% 16.000000 70.000000 6.00000 63.000000

50% 20.000000 80.000000 7.00000 75.000000

75% 24.000000 90.000000 8.00000 88.000000

max 44.000000 100.000000 10.00000 100.000000

Tutoring\_Sessions Physical\_Activity Exam\_Score

count 6607.000000 6607.000000 6607.000000

mean 1.493719 2.967610 67.235659

std 1.230570 1.031231 3.890456

min 0.000000 0.000000 55.000000

25% 1.000000 2.000000 65.000000

50% 1.000000 3.000000 67.000000

75% 2.000000 4.000000 69.000000

max 8.000000 6.000000 101.000000

Average Exam Score by Gender:

Gender

Female 67.244898

Male 67.228894

Name: Exam\_Score, dtype: float64

Number of students scoring above 85: 34

Correlation Matrix:

Exam\_Score 1.000000

Attendance 0.581072

Hours\_Studied 0.445455

Previous\_Scores 0.175079

Tutoring\_Sessions 0.156525

Physical\_Activity 0.027824

Sleep\_Hours -0.017022

Name: Exam\_Score, dtype: float64







