Analysis of Student Performance

1. Introduction

1.1 Background

This report analyzes a dataset containing academic records of 5,000 students from a private learning provider. The dataset includes various attributes such as exam scores, attendance, participation, and socio-economic factors. The objective of this analysis is to identify patterns, correlations, and factors influencing student performance.

1.2 Objectives

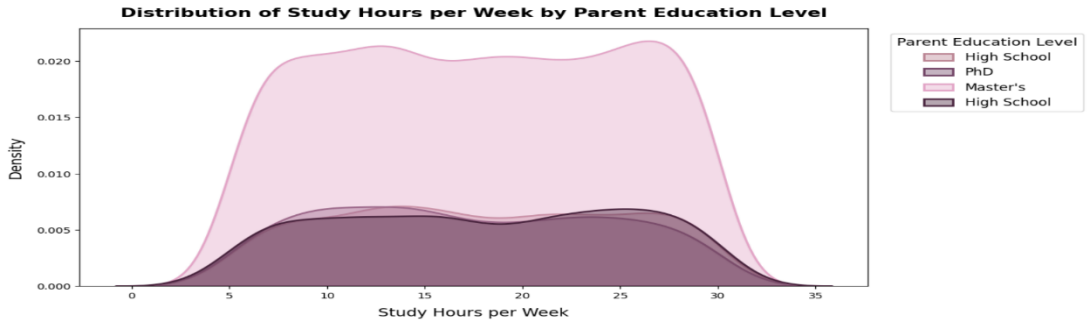
- Identify key correlations between academic performance and other attributes.
- Detect patterns in attendance, study habits, and socio-economic factors.
- Provide insights to improve student outcomes and educational strategies.

---

2. Methodology

2.1 Data Visualization

- **Heatmaps:** Correlation analysis of student performance factors.
- **Boxplots & Histograms:** Distribution analysis of grades and study-related variables.
- **Scatter Plots:** Relationships between study hours, stress levels, and total scores.



2.2 Data Preprocessing

- **Handling Missing Values:** mean Imputation for numeric missing numerical data mode for categorical data.

```
data['Assignments_Avg'].fillna(data['Assignments_Avg'].mean(),inplace=True)
data['Attendance (%)'].fillna(data['Attendance (%)'].mean(),inplace=True)
data['Parent_Education_Level'].fillna(data['Parent_Education_Level'].mode()[0],inplace=True)
```

- **Encoding Categorical Variables:** Conversion of categorical features (e.g., Gender, Extracurricular Activities) into numerical values using one-hot encoding and get dummies.

```
[36] ordinal_col=['Parent_Education_Level','Internet_Access_at_Home','Extracurricular_Activities','Family_Income_Level','Grade']
     non_ordinal_col=['Gender','Department']
```

```
[37] for col in ordinal_col:
         data[col]=le.fit_transform(data[col])
```

```
[24] for col in non_ordinal_col:
         data=pd.get_dummies(data,columns=[col],drop_first=True)
```

- **Removing Irrelevant Features**

```
[ ]  data.drop(['Student_ID','Email','First_Name','Last_Name'],axis=1,inplace=True)
```

2.3 Feature Engineering

To enhance our analysis and provide more meaningful insights, we introduced a new feature called Efficiency, which quantifies how effectively students convert study hours into performance.

We define Efficiency as the ratio of a student's total score to their study hours per week

2.4 Statistical & Correlation Analysis

- **Descriptive Statistics:** Mean, median, standard deviation for numerical features.
- **Correlation Matrix:** Identifying relationships between variables.
- **Hypothesis Testing:** Statistical tests (e.g., t-tests, ANOVA) to analyze differences in student groups.

---

3. Findings & Insights

3.1 Correlation Analysis

- **Strong Positive Correlations:**
  - **Department Math & Efficiency (r = 0.04)**: Relationship between being in the Mathematics Department and Efficiency.
  - **Family Income Level & Efficiency (r = 0.04)**: Higher Family Income is Associated with Higher Efficiency**.**
- **Negative Correlations:**
  - **Sleep Hours & Attendance (r = -0.03)**: Low sleep hours negatively impact Attendance percentage.
  - **Grade& Attendance (r = -0.57)**: Attendance increases, Grades tend to decrease.

---

4. Conclusion & Recommendations

4.1 Summary of Key Findings

- Attendance, participation, and project score impact total scores and grades.
- Stress levels and lack of sleep negatively affect academic outcomes.
- Socio-economic factors, such as parental education and income, influence performance.
- Students with Internet Access at Home Perform Better.

4.2 Recommendations for Improvement

- **Encouraging Participation:** Promoting classroom engagement through interactive learning.
- **Encourage Regular Attendance:** Monitor and improve attendance rates with engagement programs.
- **Address Stress and Sleep Deficiency**: Offer stress management programs and mental health support services.
- **Targeted Interventions:** Support for students with lower socio-economic backgrounds to bridge performance gaps.
- **Enhance Digital Learning Accessibility:** Ensure that all students have access to digital learning resources, especially at home.
- **For students aged 19-21:** Encourage better time management and stress management techniques.
- **Improve Study Techniques for Better Performance:** Since increased study hours do not necessarily translate to higher scores, students may need more effective study techniques.

---

5. Future Work

- Implement predictive modeling (e.g., regression models) to forecast student performance.

References

- Dataset used:

  https://www.kaggle.com/datasets/mahmoudelhemaly/students-grading-dataset/data