# Predicting Student Performance Based on Attendance and Socio-Economic Factors

## Introduction

Understanding student performance is critical for educators and policymakers as it helps identify the key factors that impact academic success. This study explores how attendance, parental education, study habits, and parental support influence student GPA. By applying exploratory data analysis (EDA), probability analysis, hypothesis testing, and regression modeling, we aim to determine significant predictors of student performance   
and provide insights into potential areas of improvement.

## Statistical Question

The main question of this study is: How do attendance, parental education, study hours, and parental support influence student performance, as measured by GPA? This question helps to establish a clear framework for analyzing the available data and identifying correlations and causations that may exist between these variables.

## Exploratory Data Analysis (EDA)

The dataset consists of 2,392 student records containing variables such as Study Time per Week, Absences, Parental Education, Parental Support, and GPA. Histograms generated for these variables revealed that GPA follows an approximately normal distribution, while study time and absences exhibit right-skewed distributions. This indicates that most students study relatively fewer hours per week, and a significant number have low attendance.

## Findings from Probability Analysis

- PMF Analysis: Students with high attendance (10 or fewer absences) were more likely to achieve higher GPAs compared to those with more absences. This suggests a strong correlation between attendance and performance.

- CDF Analysis: The cumulative distribution function showed that 50% of students study less than 10 hours per week, while 80% study less than 15 hours per week. This finding highlights the need for interventions that promote better study habits.

## Analytical Distribution

The GPA variable follows an approximately normal distribution but with slight skewness. This skewness may be due to external factors such as differences in teaching quality, home environments, or additional learning support that some students receive.

## Correlation and Regression Analysis

- Study Time vs. GPA: A moderate positive correlation was observed, confirming that students who study more tend to have higher GPAs.  
 - Parental Support vs. GPA: A weaker correlation was found, suggesting that while parental support is important, other factors play a larger role in student performance.  
 - Regression Model Performance:  
 - R squared = 0.918 (indicating strong predictive power)  
 - Mean Squared Error = 0.0676 (low error rate)  
 - Study Time and Parental Support positively impact GPA, while Absences negatively impact GPA.

## Hypothesis Testing

A t-test was conducted to determine whether attendance significantly affects GPA. The results showed a p-value of 0.0, confirming that students with high attendance perform significantly better academically.

## Conclusion

This study highlights that attendance and study habits are the strongest predictors of academic performance. Parental support has a smaller yet positive effect. The findings suggest that encouraging better attendance and promoting effective study habits can significantly improve student outcomes. Future research should incorporate additional socio-economic factors and explore non-linear models to gain a more comprehensive understanding of student success.