📊 Student Performance & Behavior Dataset

Introduction

This report presents an analysis of the Student Performance & Behavior Dataset, aiming to uncover insights into factors affecting student academic outcomes. The dataset contains information on 5,000 students from a private educational institution, including demographics, academic performance metrics, and behavioral factors. The goal is to identify key factors influencing student success and suggest data-driven recommendations for potential academic improvement strategies.

Methodology

The analysis was conducted in the following stages:

1. **Data Preprocessing:**

   - Loaded the dataset into a Pandas Data Frame.
   - Dropped irrelevant columns (Student ID, names, email).
   - Analyzed categorical and numerical features to understand data characteristics and potential issues.

2. **Handling Missing Values:**

   - **Categorical Columns:**
     - Identified a high percentage of missing values (35.88%) in the 'Parent_Education_Level' column. Due to the low correlation with other features, the column was dropped to avoid introducing bias through imputation.

- **Numerical Columns:**
    - Analyzed correlation and distributions of 'Attendance (%)' and 'Assignments_Avg' columns, which contained missing values.
    - Applied Regression Imputation to estimate missing values for 'Attendance (%)'.
    - Applied Random Imputation to estimate missing values for 'Assignments_Avg'.

3. Outlier Detection and Handling:

- Used box plots to visually identify potential outliers in numerical features.
- Employed the IQR method to detect and replace numerical outliers with reasonable values.
- Applied the threshold method to address rare categorical values, ensuring that categorical outliers were replaced with the mode.

4. Data Exploration and Visualization:

- Assessed data skewness using histograms and KDE plots for numerical features.
- Created count plots to analyze categorical features and their relationship with grades.
- Generated a correlation matrix to identify relationships between numerical features.
- Employed a pair plot to further investigate the combined impact of multiple variables on student performance.

- Developed line plots to analyze trends in academic performance based on study hours, attendance, and stress levels.
- Created bar plots to visualize the impact of extracurricular activities and family income on grades.

5. **Feature Engineering:**
   - Binned variables like Study Hours, Attendance, and Stress level to help understand the effect of these factors within certain ranges.
   - Calculated average scores within these bins to identify trends.

Findings

- **Key Predictors of Academic Performance:**
  - Attendance, assignments average, midterm score, final score, quizzes average, and projects score emerged as significant predictors of total score and student grades.
- **Impact of Study Hours:**
  - A positive trend was observed between study hours and final scores, with potential diminishing returns beyond a certain threshold.
- **Influence of Attendance:**
  - A negative relationship was initially observed between lower attendance (70-80%) and scores, but this relationship shifted positively for higher attendance (80-90% and above).
- **Stress Level Correlation:**

- A negative trend was observed between stress levels and final scores, suggesting that stress can negatively impact academic performance.
- **Extracurricular Activities:**
  - Students with higher grades tended to have lower participation rates in extracurricular activities compared to students with lower grades, potentially due to time management or academic focus.
- **Socioeconomic Factors:**
  - Family income level appeared to have a positive correlation with grades, suggesting potential socioeconomic influences on student achievement.

Conclusion

The analysis of the Student Performance & Behavior Dataset provides valuable insights into factors affecting student success. Key findings suggest that attendance, consistency in assignments, and efficient study habits are crucial for academic performance. Stress levels appear to have a negative impact, highlighting the need for student support in this area. Socioeconomic background also seems to influence academic outcomes. By understanding these trends, institutions can make informed decisions to improve teaching strategies, allocate resources effectively, and foster an environment that supports student growth and achievement.