**Executive Summary**

Our project sought to identify critical factors predicting high school students' academic performance. Using a dataset comprising information on demographics, study habits, parental involvement, and extracurricular activities for 2,392 students, we applied various data mining techniques to build a predictive model.

**Key Findings**

1. **Study Time and Academic Performance:**
   * Higher weekly study time is positively correlated with better academic performance. Students who spend more hours studying each week tend to achieve higher grades.
2. **Parental Support:**
   * The level of parental support plays a significant role in predicting academic success. Students with moderate to very high parental support levels are likelier to perform well academically.
3. **Extracurricular Activities:**
   * Participation in extracurricular activities, particularly sports and music, is associated with higher academic performance. These activities contribute positively to students' grades.
4. **Demographic Factors:**
   * Parental education level and ethnicity are significant predictors of academic outcomes. Students whose parents have higher education levels tend to achieve better grades. There are also observable differences in educational performance across different ethnic groups.
5. **Absences and Tutoring:**
   * Higher numbers of absences negatively impact academic performance, while receiving tutoring has a positive effect. Students with fewer absences and those who receive tutoring are more likely to achieve higher grades.

**Conclusion**

Our predictive model successfully classifies students' grade classes based on demographic details, study habits, parental involvement, and extracurricular activities. The findings highlight the importance of a holistic approach to education, where various aspects of a student's life are considered to support their academic success. Moving forward, these insights can inform policy decisions and educational strategies to improve student outcomes.

**Logistic Regression**

**Key Predictors:**

1. Weekly study time
2. Parental support
3. Participation in sports

**Insights:**

* Increased study time and parental support are associated with higher grades.
* Participation in sports positively influences academic outcomes.

**Decision Tree**

**Key Predictors:**

1. Study time
2. Absences
3. Parental support

**Insights:**

* Absences negatively impact academic performance.
* High study time and parental support remain significant positive factors.

**K-Nearest Neighbors (KNN)**

**Key Predictors:**

1. Study time
2. Parental support
3. Extracurricular activities

**Insights:**

* Higher study time and parental support are linked to better academic performance.
* Participation in extracurricular activities is also a positive factor.

**Random Forest**

**Key Predictors:**

1. Study time
2. Parental support
3. Extracurricular activities

**Insights:**

* Increased study time and parental support are linked to better academic performance.
* Participation in extracurricular activities contributes positively to students' grades.

**Overall Findings**

Across all models, certain variables consistently emerged as significant predictors of academic performance:

1. **Study Time**: Higher weekly study time is consistently linked to better grades.
2. **Parental Support**: Greater parental involvement and support are crucial for academic success.
3. **Extracurricular Activities**: Participation in sports and other extracurricular activities positively influences grades.
4. **Absences**: Higher numbers of absences have a detrimental effect on academic performance.

**Conclusion**

The consistency in findings across different models underscores the importance of study habits, parental involvement, and extracurricular activities in predicting academic performance. These insights provide a robust foundation for developing targeted interventions and support systems to enhance student outcomes. Educators and policymakers can use these findings to implement strategies that foster a supportive and enriching educational environment.

**Plot Interpretations**

1. **Study Time Weekly vs. GPA**
   * The scatter plot shows the relationship between weekly study time and GPA.
   * Observations:
     + A slight positive trend suggests that increased study time generally correlates with a higher GPA.
     + However, the scatter is quite dispersed, indicating that other factors significantly influence GPA.
2. **Age vs GPA**
   * The plot displays the distribution of GPA across different ages.
   * Observations:
     + The GPA distribution appears relatively consistent across age groups (15, 16, 17, and 18 years old).
     + There is no clear age-related trend in GPA, suggesting that age alone is not a strong predictor of academic performance.
3. **Gender vs GPA**
   * This box plot compares GPA distributions between genders.
   * Observations:
     + The medians for both genders are similar, indicating no significant difference in GPA based on gender.
     + The interquartile ranges and overall distribution also appear similar, reinforcing the lack of gender disparity in academic performance.