BrightPath Academy:
Student Performance Analysis Report

Completed By:

Thapelo Mogashoa   **600298**

Janco Nieuwoudt   **600175**

Dean Jacobus Andreas van Zyl   **600367**

Executive Summary

This report presents the findings of a comprehensive data analysis study conducted on student performance data at BrightPath Academy. The analysis implemented machine learning approaches to recognize performance-influencing variables in students and to create a model which predicts academic risks among students before major deteriorations occur.

Based on our modelling analysis using several different models, **absences from class were consistently identified as the most significant predictor of student GPA and overall grade class scores**. This highlights the critical importance of regular attendance for academic success at BrightPath Academy.

Academic performance at BrightPath Academy demonstrated significant correlation between parental support together with tutoring participation and study habits but student demographics played a minimal role. The discovered data streamlines meaningful solutions through which educators might create specific intervention strategies for supporting student achievement.

Key Findings

1. Attendance Is Critical for Academic Success
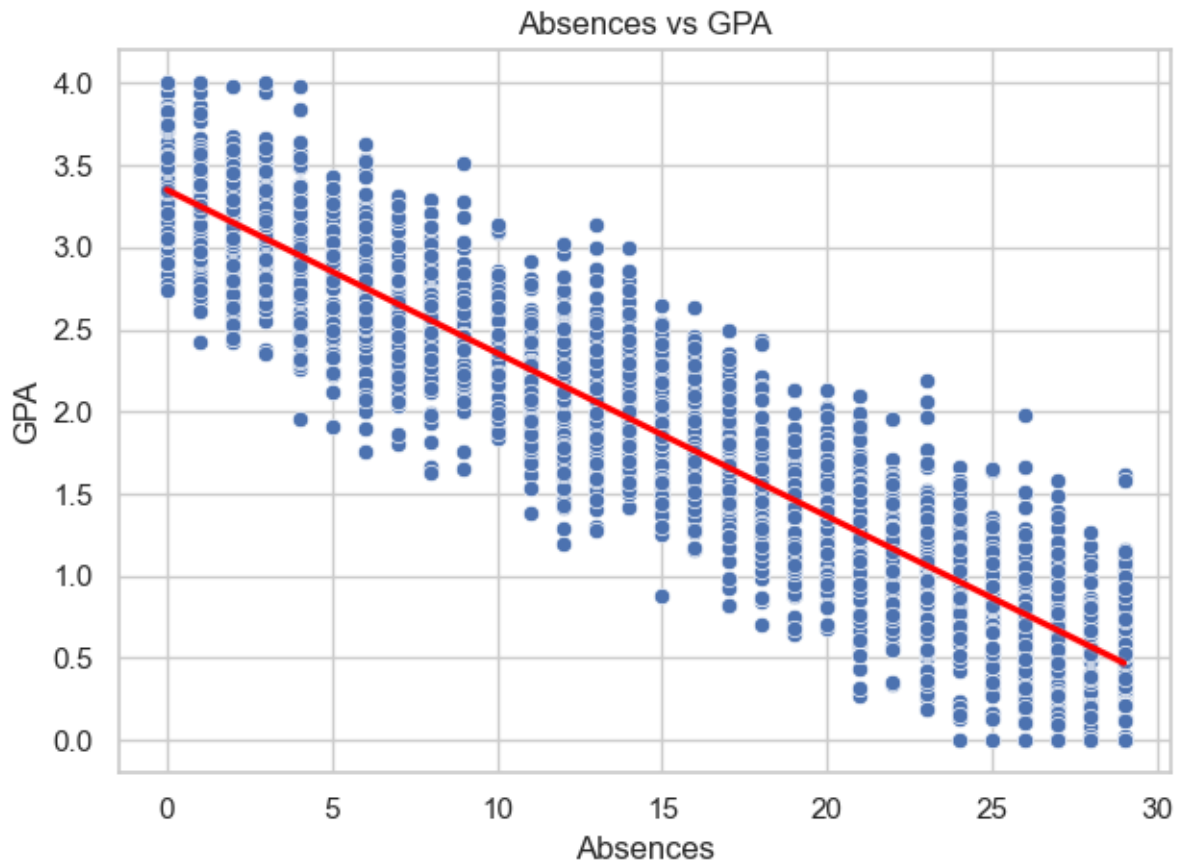
The data reveals a strong negative correlation (r = 0.73) between absences and GPA. As shown in the scatter plot with regression line, there is a clear linear relationship: as the number of absences increases, GPA consistently decreases. Students with fewer than 5 absences typically maintain GPAs above 3.0, while those with more than 20 absences frequently fall below a 2.0 GPA.

This finding suggests that attendance monitoring should be a primary focus for early intervention strategies, as it serves as the strongest predictor of academic performance.

## 2. Study Time Shows Positive but Limited Impact

Study Time Weekly vs GPA
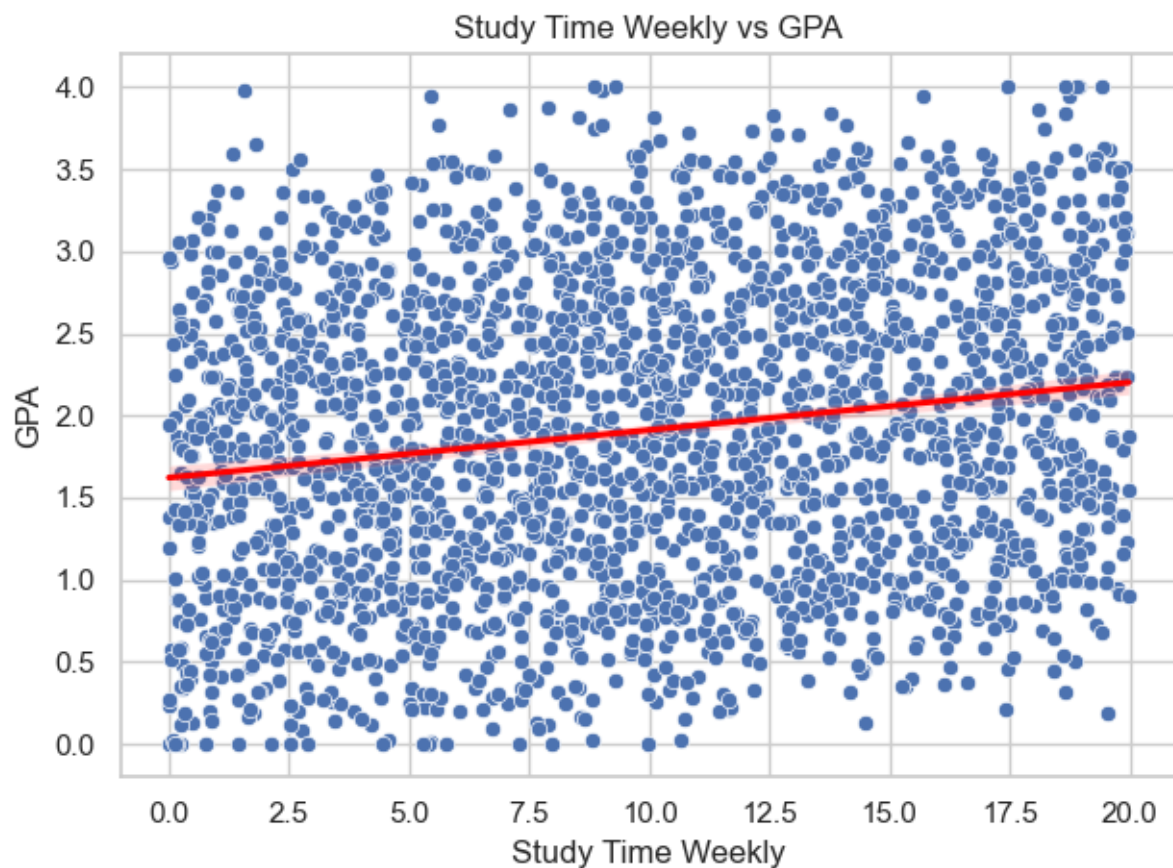
The relationship between weekly study time and GPA shows a positive but relatively weak correlation. The scatter plot with regression line indicates a slight upward trend, suggesting that while increased study time is beneficial, its impact is moderated by other factors.

This finding suggests that simply encouraging students to study more may not be sufficient; the quality and effectiveness of study habits may be more important than quantity alone.

# 3. Parental Support Significantly Influences Academic Performance
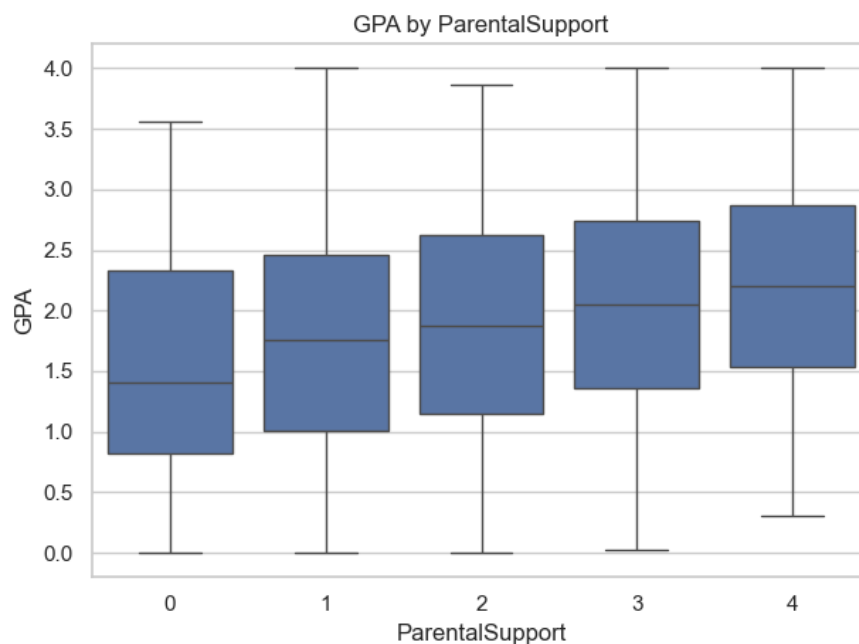
The box plot of GPA by parental support level shows a clear trend: as parental support increases from level 0 (none) to level 4 (very high), the median GPA steadily rises. Students with the highest level of parental support (level 4) have median GPAs above 2.0, while those with no parental support (level 0) have median GPAs below 1.5.
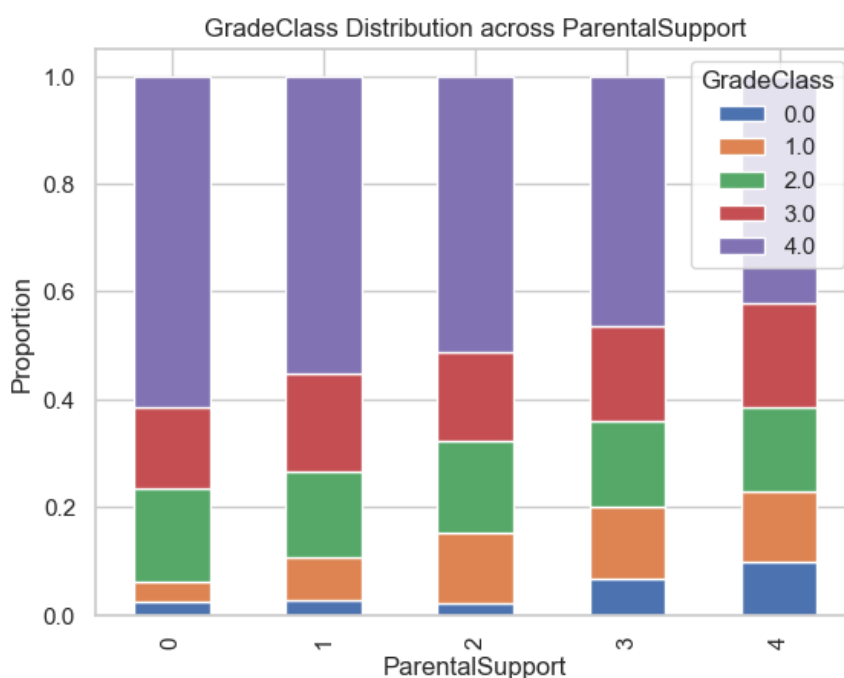

This stacked bar chart further illustrates the impact of parental support on grade distribution. The proportion of 'A' and 'B' grades (blue and orange sections) increases with higher parental support levels, while the proportion of 'F' grades (purple sections) decreases significantly.

4. Demographic Variables Have Minimal Impact on Performance


Correlation Heatmap
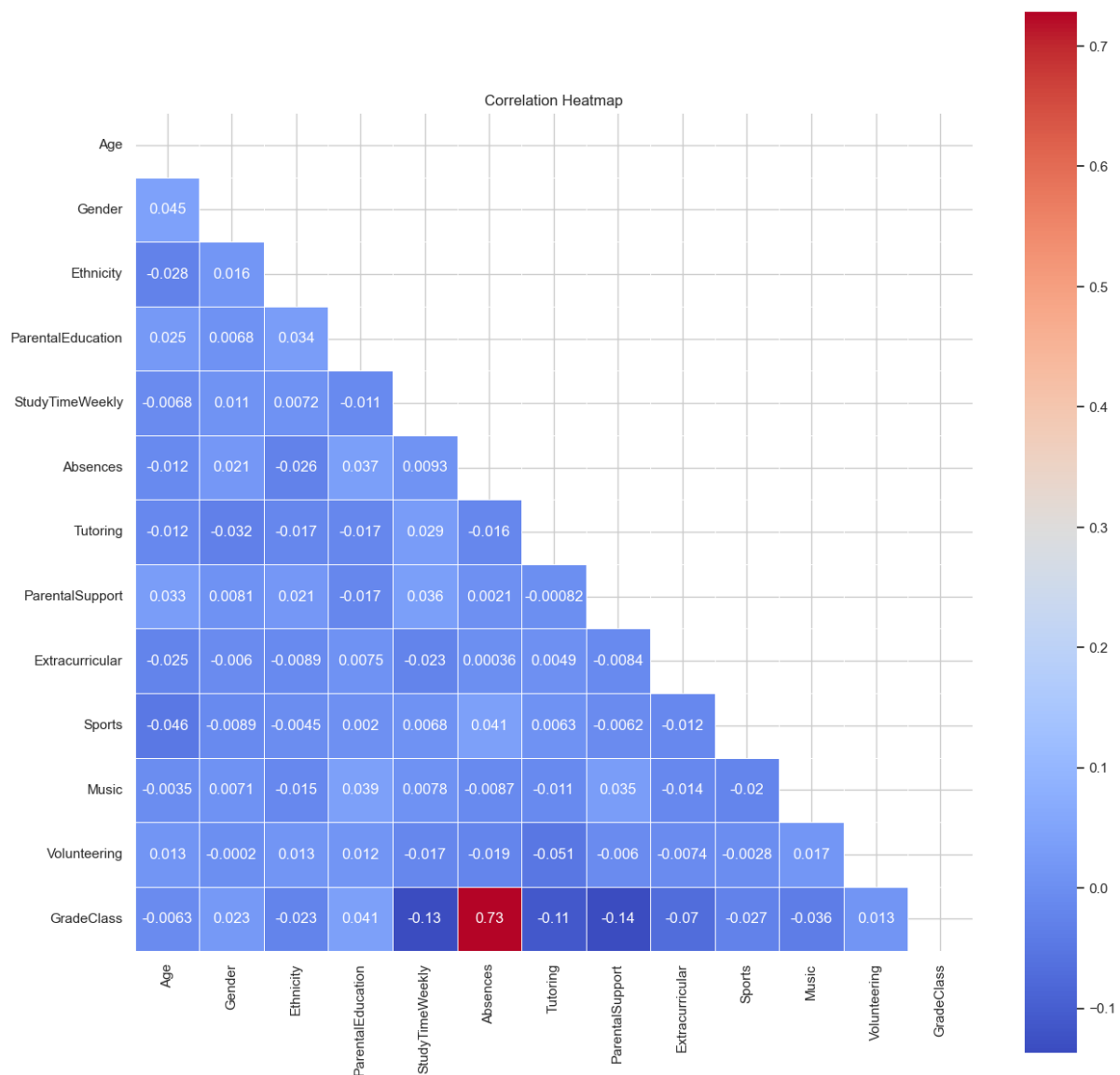
The correlation heatmap (Image 3) shows that demographic factors such as age, gender, and ethnicity have very weak correlations with GPA and grade class, with correlation coefficients ranging from -0.05 to 0.05. This suggests that academic performance is more strongly influenced by behavioural and environmental factors than by demographic characteristics.

5. Significant Class Imbalance in Grade Distribution

Boxplot of Age

Boxplot of StudyTimeWeekly

Boxplot of Absences

Boxplot of GPA
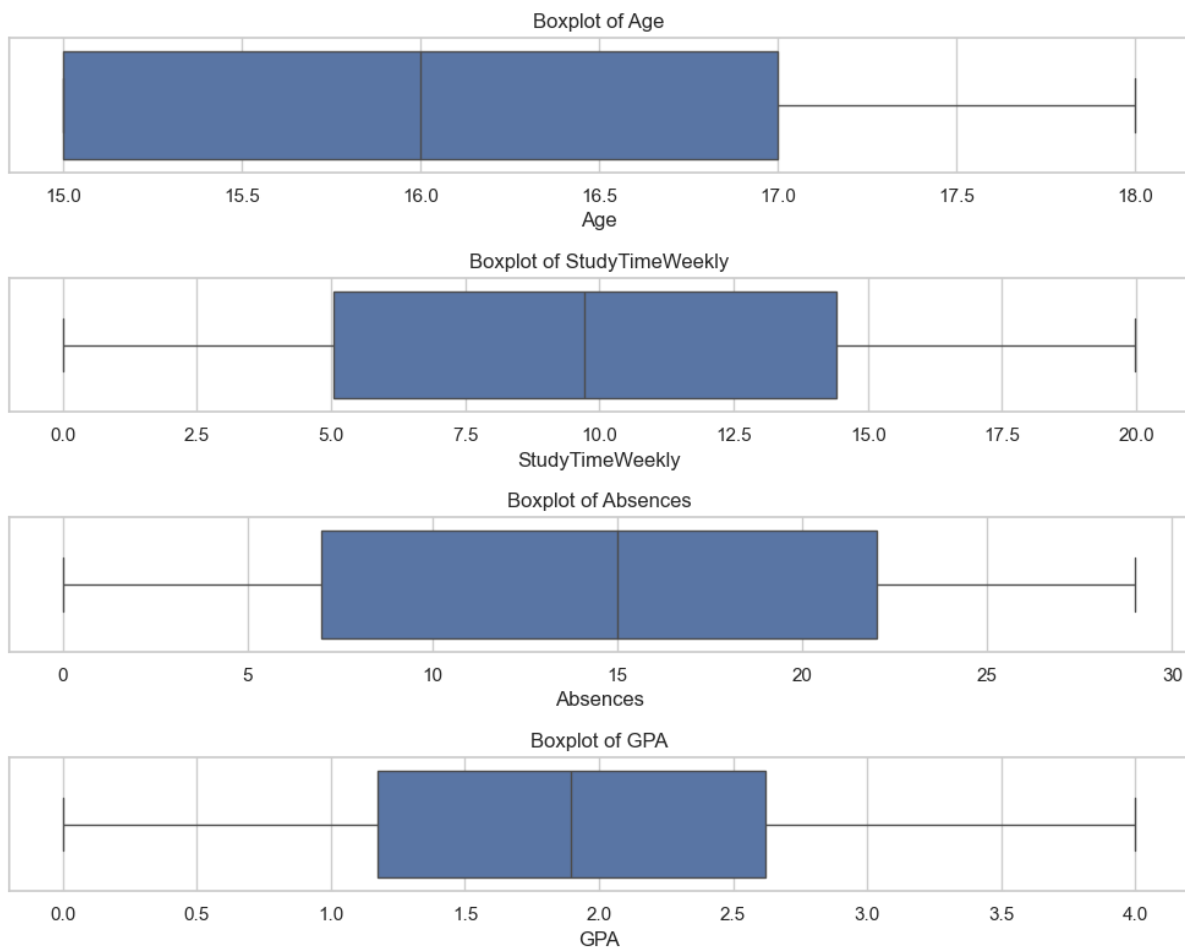
The grade class distribution shows a significant imbalance, with approximately 50% of students falling into the 'F' category. This suggests that a large portion of the student body is struggling academically and may benefit from targeted interventions.

6. Distribution of Key Academic Factors

GradeClass Distribution
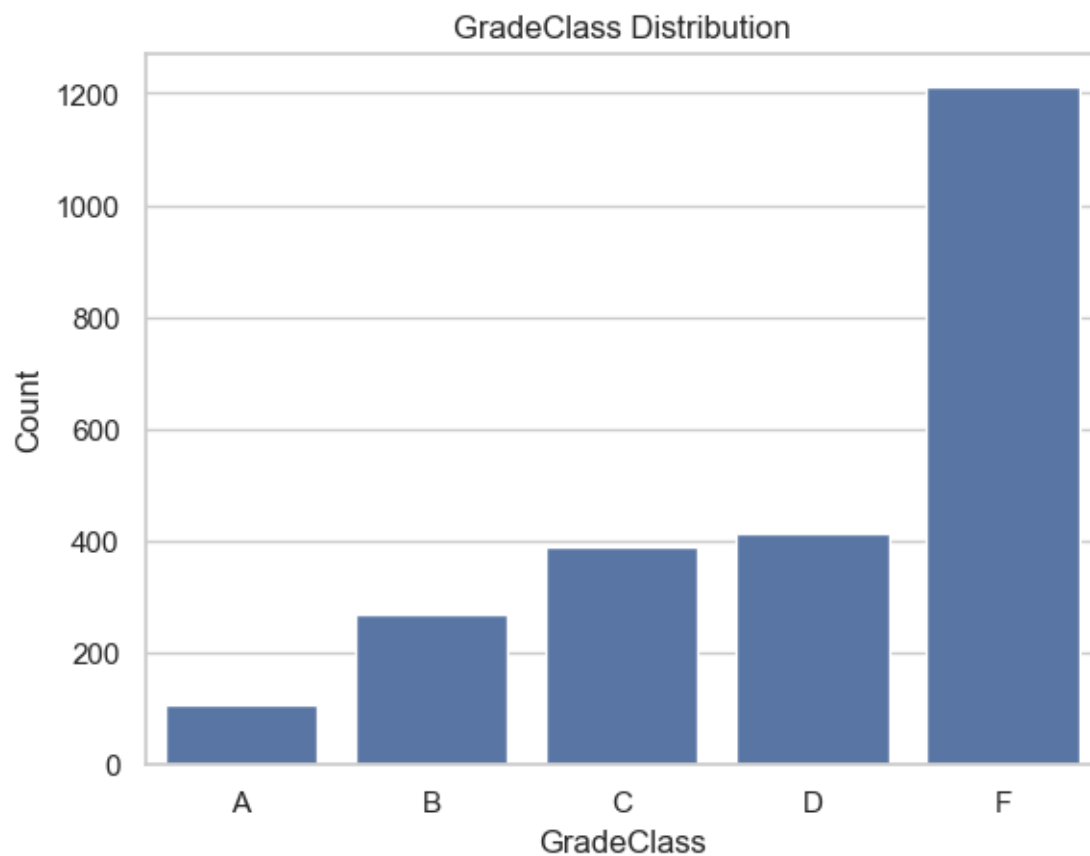
The boxplots show the distributions of key variables:

- Age: Most students are between 15-17 years old

- Study Time: Weekly study hours vary widely, typically between 5-15 hours

- Absences: Most students have between 7-20 absences per year

- GPA: The distribution is relatively uniform across the 0-4.0 range, with a slight concentration around 1.5-2.5

7. GPA Distribution Patterns
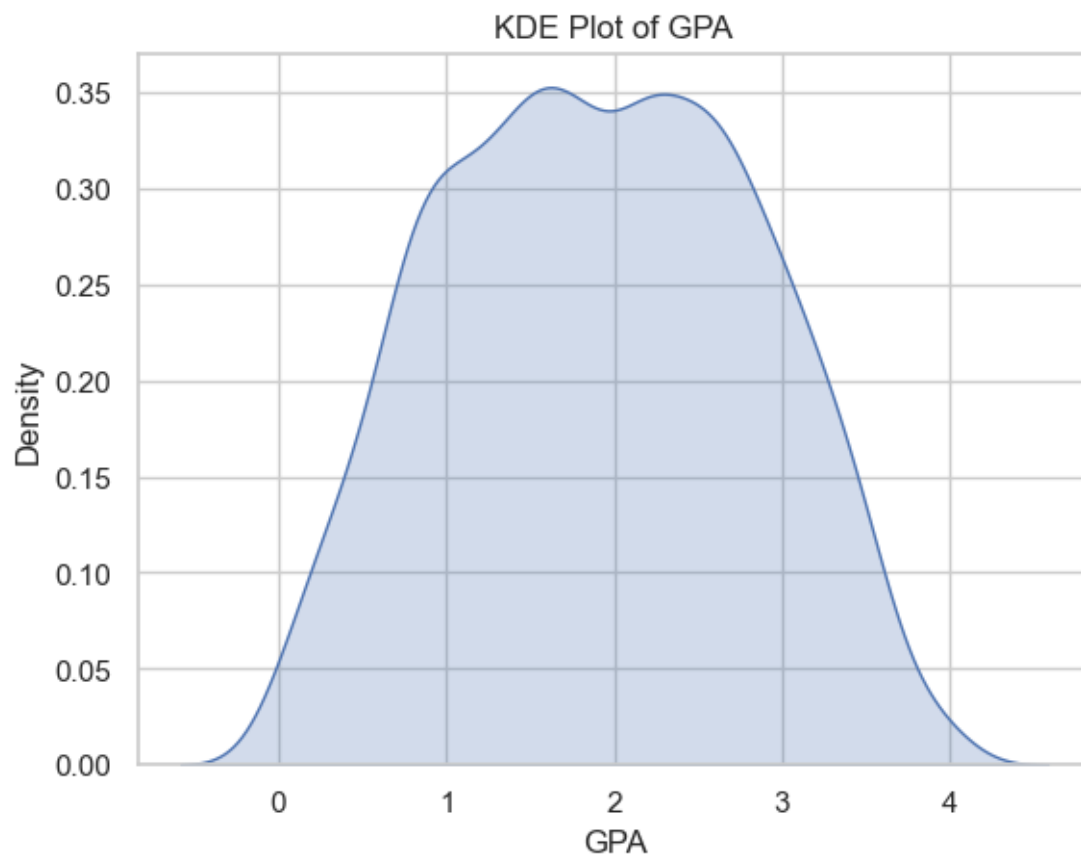
The KDE plot shows that GPA distribution has a relatively flat shape with slightly higher density in the middle range (1.5-2.5), indicating a broad spread of academic performance levels among students.

Model Development

During our model development we decided to train 5 models to see which give us the highest score. The models we trained are the following:

Model	Accuracy Score (%)
Random Forest Classifier	74.94
K Neighbours Classifier	63.25
Logistic Regression	80.79
Gradient Boosting Classifier	76.40
SVC	77.24
Sequential Deep Learning	77.66

As you can see from this table, the Logistic Regression model was the best for this use case. From there we evaluated and did a cross comparison between the two highest performing models (non-deep learning). After comparing logistic regression and SVC and tuning their parameters, Logistic Regression achieved the highest performing score at 82.02%.
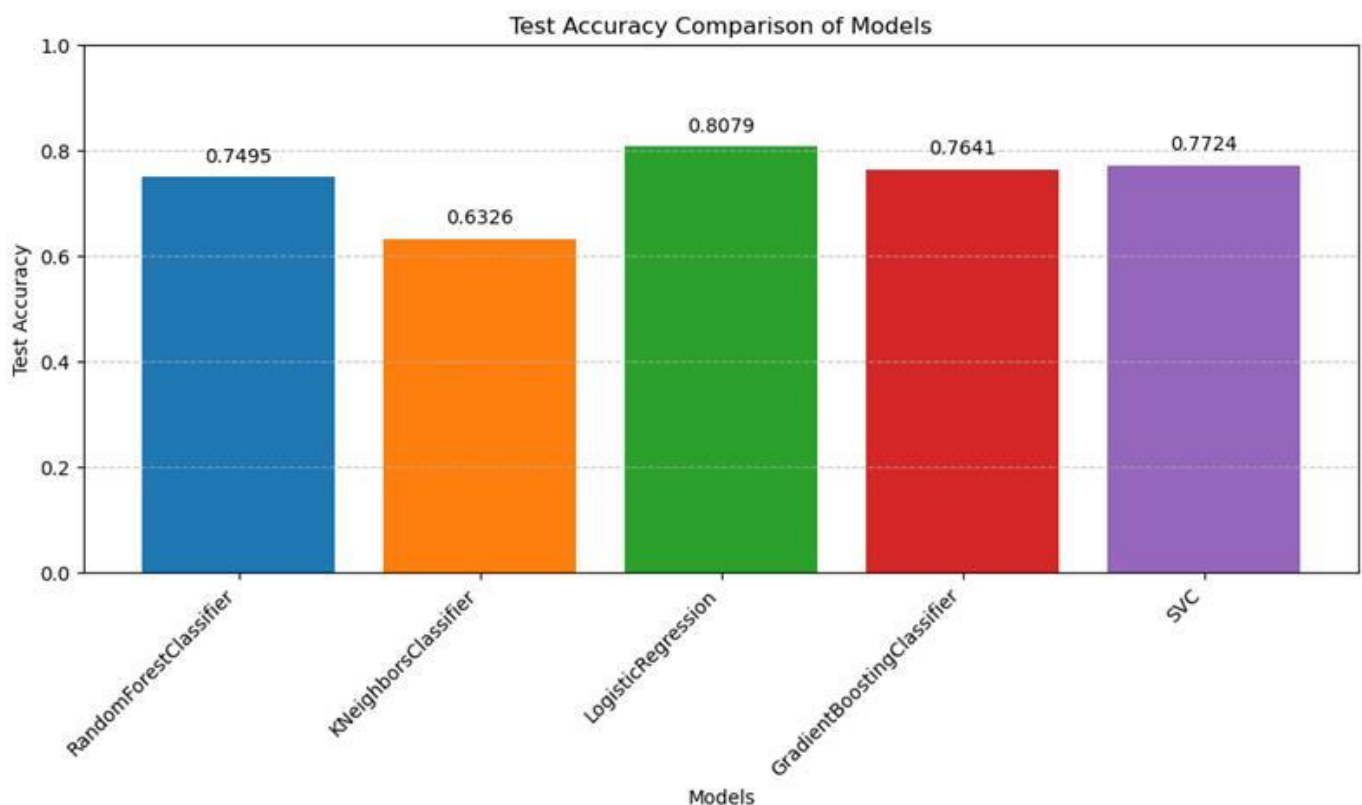
Machine Learning Insights

Our predictive modelling identified the following factors as the most significant predictors of student academic performance (in order of importance):

1. **Absences**: By far the strongest predictor

2. **Parental Support**: Significant positive impact

3. **Tutoring Participation**: Moderate positive impact

4. **Study Efficiency**: Study time relative to absences

5. **Extracurricular Activity Participation**: Slight positive impact

Recommendations

Based on the analysis, we recommend the following strategies to improve student academic performance at BrightPath Academy:

1. **Implement an Early Attendance Warning System**: Develop an automated system to flag students with increasing absence patterns before they reach critical levels. Our data suggests intervention should begin after 5-7 absences.

2. **Strengthen Parental Engagement Programs**: Create targeted initiatives to increase parental involvement, particularly for students with currently low levels of parental support. Consider parent education workshops, regular communication channels, and parent-teacher collaboration opportunities.

3. **Expand Tutoring Access**: Increase availability and promotion of tutoring services, with priority given to students with higher absence rates or lower parental support levels.

4. **Develop Personalized Study Strategy Support**: Rather than simply encouraging more study time, provide guidance on effective study techniques tailored to individual learning styles and subjects.

5. **Balance Academic and Extracurricular Activities**: Continue to promote extracurricular participation while ensuring it doesn't detract from academic commitments.

6. **Create a Comprehensive Dashboard**: Implement a real-time monitoring system that integrates attendance, academic performance, and participation metrics to enable proactive intervention.

Conclusion

The data clearly demonstrates that attendance is the single most important factor affecting student performance at BrightPath Academy. By focusing intervention efforts on improving attendance and strengthening parental support, the school can make significant strides in raising overall academic achievement.

The predictive model developed through this analysis provides a powerful tool for identifying at-risk students early, allowing for targeted interventions before academic performance severely deteriorates. With systematic implementation of the recommended strategies, BrightPath Academy is well-positioned to fulfil its mission of empowering all students to reach their full potential.

Additional Information:

Here is our GitHub repository link:
https://github.com/CantCode29/MLG382ProjectFolder

Link back to the Render App:

https://mlg382projectfolder.onrender.com