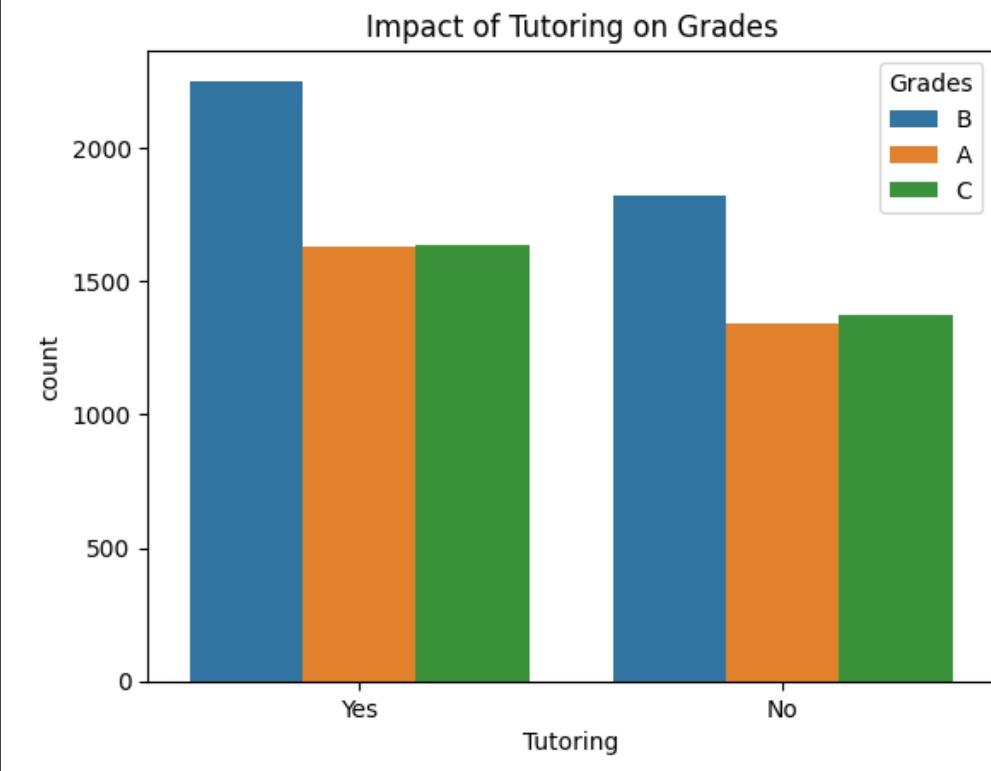
**Factors affecting university student grades**

Prepared by: Abdelrahman Harb

Prepared for: Rafi Ptashny, Chief Data Analyst

Date: 9 feb 2025

**Data Analysis:**



1. **Tutoring and Grade Distribution:**

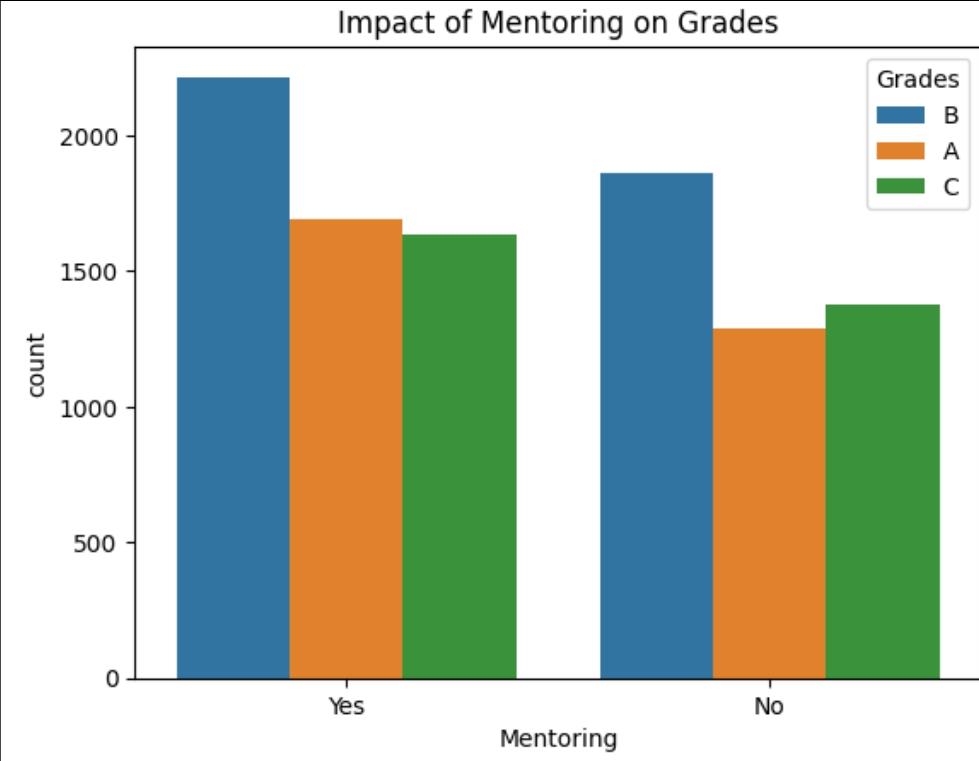
* "**Yes**" Tutoring: Among students who received tutoring, the most common grade is "B," followed by "A," and then "C."
* "**No**" Tutoring: Among students who did not receive tutoring, the distribution is slightly different. "B" is still the most frequent, but the number of "A" grades is notably lower, and the number of "C" grades is slightly higher compared to the "Yes" tutoring group.

1. **Impact of Tutoring:**

* Increased "**A**" Grades: A higher proportion of students who received tutoring achieved an "A" grade compared to those who did not. This suggests that tutoring might have a positive impact on achieving higher grades.
* Reduced "**C**" Grades: There's a slight decrease in the proportion of "C" grades among students who received tutoring.
* "**B**" Grades Remain High: While tutoring seems to have a positive influence on "A" grades, the number of "B" grades remains relatively high in both groups. This could indicate that other factors also play a significant role in achieving a "B" grade.

1. **Potential Implications:**

* Tutoring Effectiveness: The data suggests that tutoring could be an effective intervention for improving student grades, particularly for those aiming for an "A.”
* **Program Evaluation:** This data supports the idea that tutoring programs can be beneficial. InvestInMinds could use this information to evaluate and potentially expand their tutoring initiatives.
* **Targeted Support:** The findings suggest that tutoring might be particularly helpful for students aiming for higher grades ("A").
* **Further Research:** InvestInMinds could conduct more rigorous research (potentially including randomized controlled trials) to establish the causal impact of tutoring and to identify other factors that contribute to student success. They might also want to explore why the "B" grade is so prevalent even with tutoring, which could point to areas for additional intervention.



1. **Grade Distribution by Mentoring Status:**

* "**Yes**" Mentoring: Among students who had mentors, the most frequent grade is "B," followed by "A," and then "C."
* "**No**" Mentoring: The distribution is similar for students without mentors, with "B" being the most common, followed by "A," and then "C."

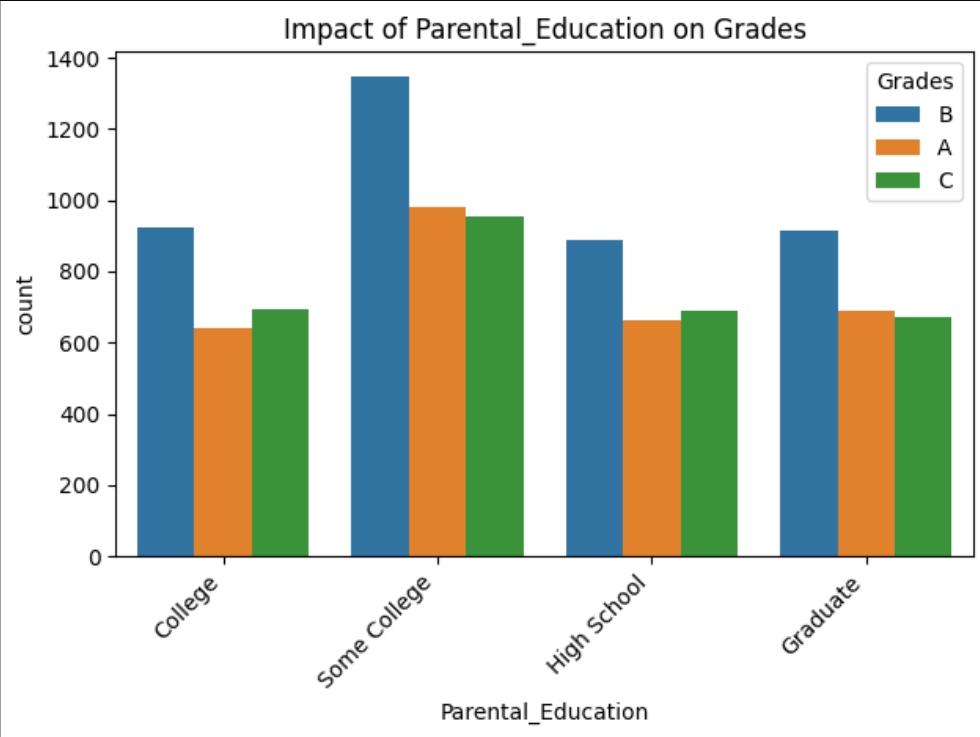
1. **Impact of Mentoring:**

* Slightly Higher "**A**"s with Mentoring: A slightly higher number of students who had mentors achieved an "A" grade compared to those who didn't have mentors. This suggests a potential positive influence of mentoring on achieving higher grades.
* Similar "**B**" and "**C**" Trends: The counts of "B" and "C" grades are relatively similar between the two groups (mentored vs. not mentored).

1. **Potential Implications:**

* Mentoring's Potential Benefit: The data hints at a possible positive relationship between mentoring and achieving higher grades, particularly "A"s.

**In summary**, the chart suggests a possible positive association between mentoring and higher grades, but further research is needed to confirm a causal relationship and rule out the influence of other factors. The information can be valuable for InvestInMinds in evaluating and refining their programs.



1. **Grade Distribution Across Parental Education Levels:**

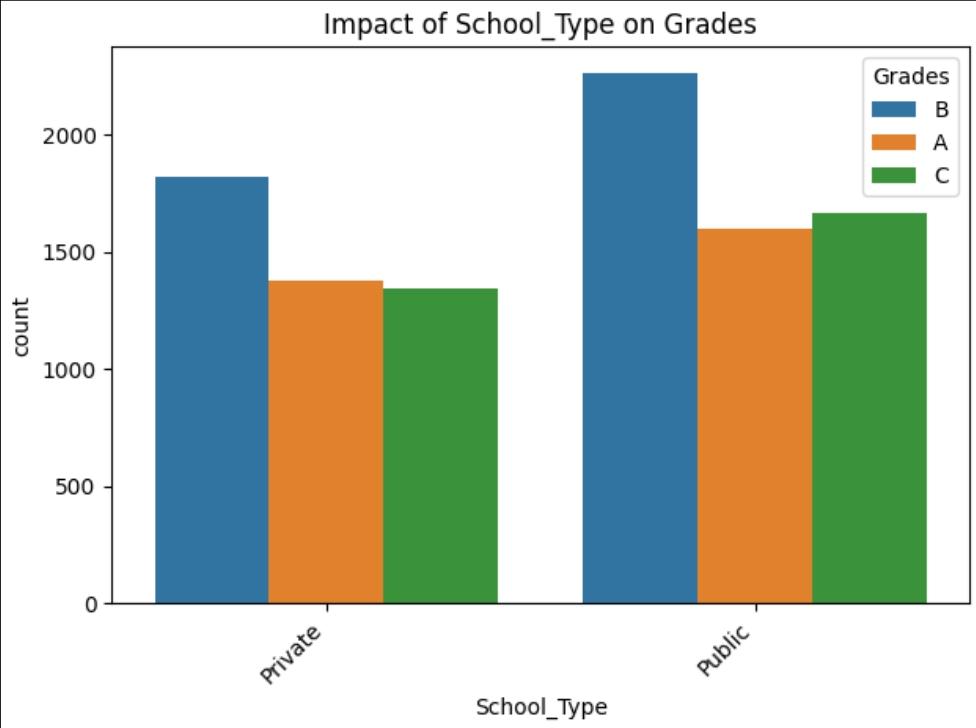
* "**College**": Among students whose parents' highest education is "College," the most frequent grade is "B," followed by "A," and then "C."
* "**Some College**": The trend is similar for students whose parents attended "Some College," with "B" being the most frequent, followed by "A," and then "C." However, the overall count in this category is the highest.
* "**High School**": For students whose parents' highest education is "High School," the distribution is slightly different. The count of "B" grades is still the highest, but the count of "C" grades is relatively higher compared to the "College" and "Some College" categories.
* "**Graduate**": Students whose parents are "Graduates" show a similar pattern to the "College" and "Some College" groups, with "B" being the most common, followed by "A," and then "C."

1. **Impact of Parental Education on Grades:**

* Higher Parental Education, Potentially Higher "**A**"s: There is a tendency, though not absolute, for the proportion of "A" grades to be slightly higher in the "College," "Some College," and "Graduate" categories compared to the "High School" category. This suggests that higher parental education might be associated with a slightly increased likelihood of students achieving higher grades.
* "**B**" Grade Consistency: The "B" grade remains the most frequent across all parental education levels.
* "**C**" Grades and Lower Parental Education: The "High School" category shows a slightly higher proportion of "C" grades compared to the other categories, suggesting that students whose parents have a high school education might be slightly more likely to receive lower grades.

1. **Potential Implications:**

* Parental Education as a Factor: The data hints at parental education being a potential factor influencing student grades.
* **Targeted Interventions:** The data suggests that students whose parents have lower levels of education might benefit from additional support and resources.
* **Holistic Approach:** InvestInMinds should consider a holistic approach that addresses multiple factors, recognizing that parental education is just one piece of the puzzle.



1. **Grade Distribution Across School Types:**

* "**Private**": Among students attending private schools, the most frequent grade is "B," followed by "A," and then "C."
* "**Public**": The distribution is similar for students attending public schools, with "B" being the most common, followed by "A," and then "C."

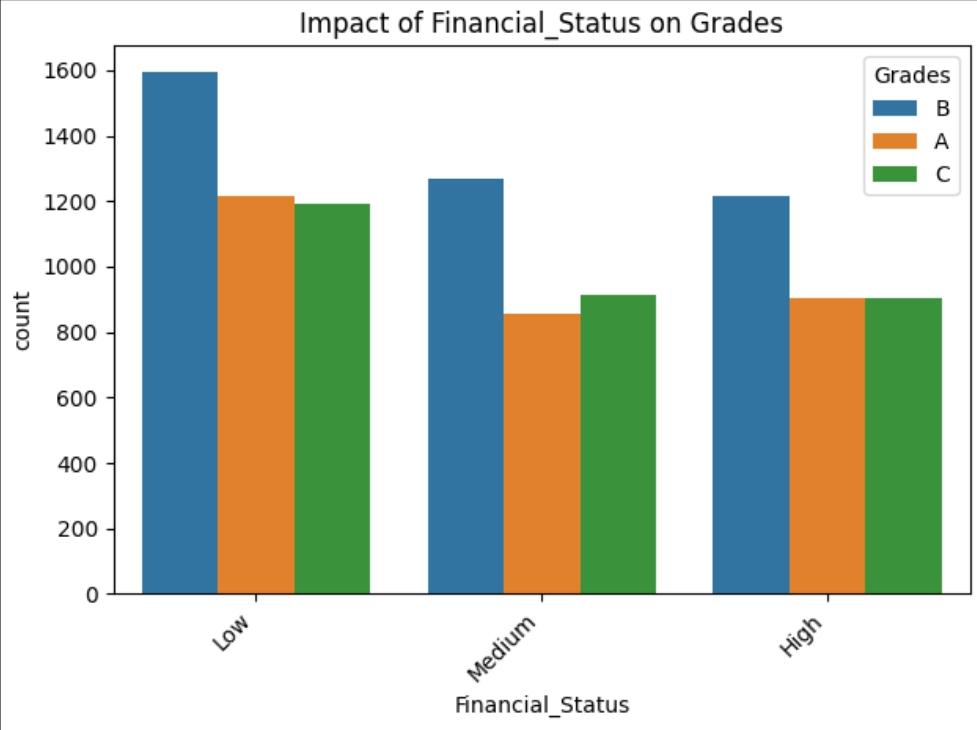
1. **Impact of School Type on Grades:**

* **Similar Trends:** The overall patterns of grade distribution are quite similar between private and public schools. Both show a higher concentration of "B" grades, followed by "A" and then "C."
* **Slight Variations:** There are some minor variations in the heights of the bars, suggesting slightly different proportions of grades between the two school types. However, these differences don't appear to be substantial at first glance.

1. **Potential Implications:**

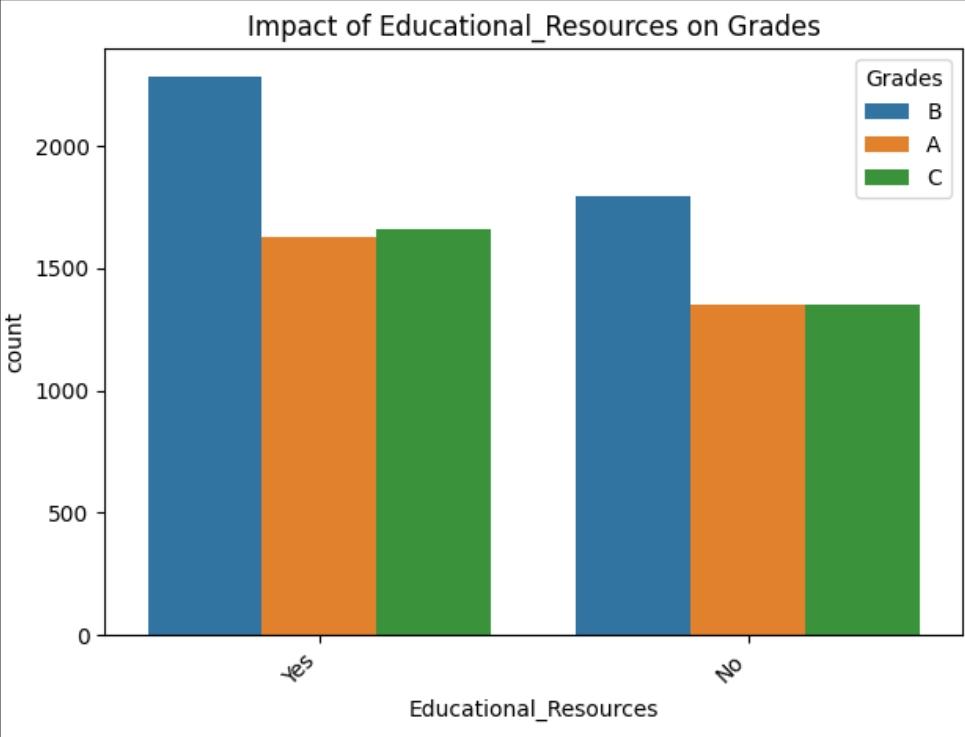
*School Type Might Not Be the Dominant Factor:* The data suggests that the type of school (private or public) might not be the most influential factor in determining student grades, as the distributions are relatively similar.

**In summary**, the chart suggests that school type alone might not be a strong predictor of student grades, and other factors are likely playing a more significant role. InvestInMinds should consider this when developing strategies to support student success



* **Holistic Understanding:** InvestInMinds should consider a holistic understanding of student success, recognizing that financial status is just one factor among many.
* **Focus on Multifaceted Support:** Given the relatively consistent grade distributions, InvestInMinds might want to focus on providing multifaceted support that addresses various needs, such as academic resources, mentoring, tutoring, and socio-emotional support, rather than solely focusing on financial assistance.
* **Further Research:** InvestInMinds could conduct further research to explore how financial status interacts with other factors to affect student grades. They might also want to investigate the specific challenges faced by students from different financial backgrounds and how these challenges can be addressed through targeted interventions.

**In summary**, the chart suggests that financial status alone might not be a strong predictor of student grades, and other factors are likely playing a more significant role. InvestInMinds should consider this when developing strategies to support student success. They should focus on a wide range of support mechanisms and conduct further research to gain a deeper understanding of the complex factors influencing student outcomes.



1. **Grade Distribution Across Resource Availability:**

* "**Yes**" Resources: Among students with access to educational resources, the most frequent grade is "B," followed by "A," and then "C."
* "**No**" Resources: The trend is similar for students without access to resources, with "B" being the most common, followed by "A," and then "C."

1. **Impact of Educational Resources on Grades:**

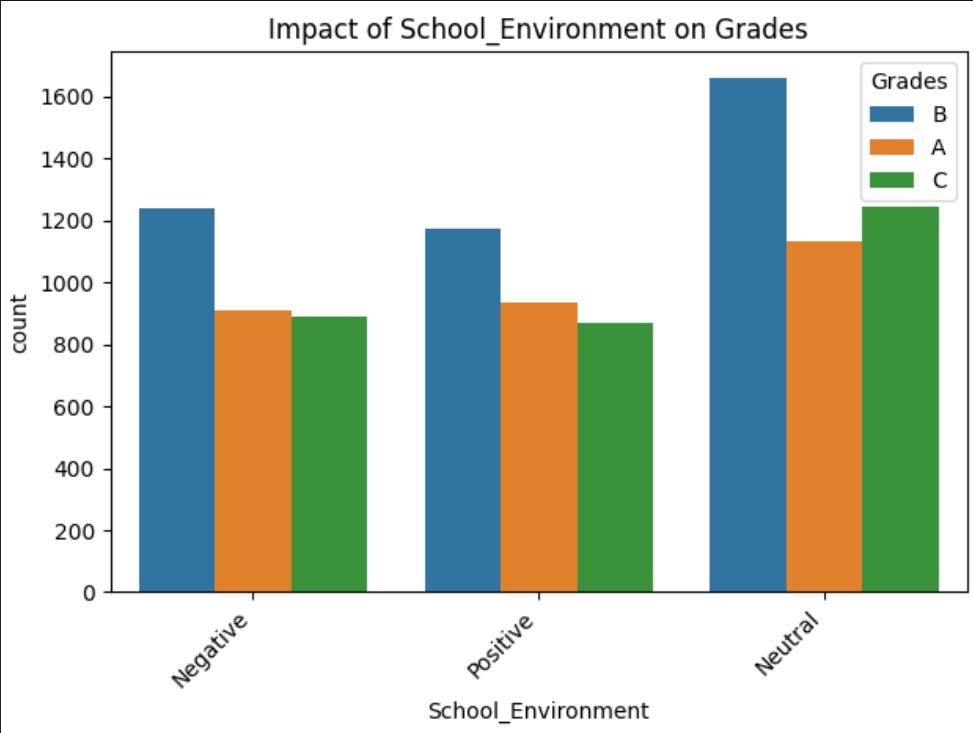
* Slightly Higher "**A**"s with Resources: There appears to be a slightly higher proportion of "A" grades among students who have access to educational resources compared to those who don't.
* Similar "**B**" and "**C**" Trends: The counts of "B" and "C" grades are relatively similar between the two groups (resource access vs. no access).

1. **Potential Implications:**

* **Resources Might Have a Positive Influence:** The data hints at a possible positive relationship between access to educational resources and achieving higher grades, particularly "A"s.

**Resource Provision as a Potential Strategy:** The data suggests that providing educational resources could be a beneficial strategy for improving student outcomes.

**Targeted Resource Allocation:** InvestInMinds might consider focusing on providing resources to students who lack access, particularly those who are aiming for higher grades.



1. **Grade Distribution Across School Environments:**

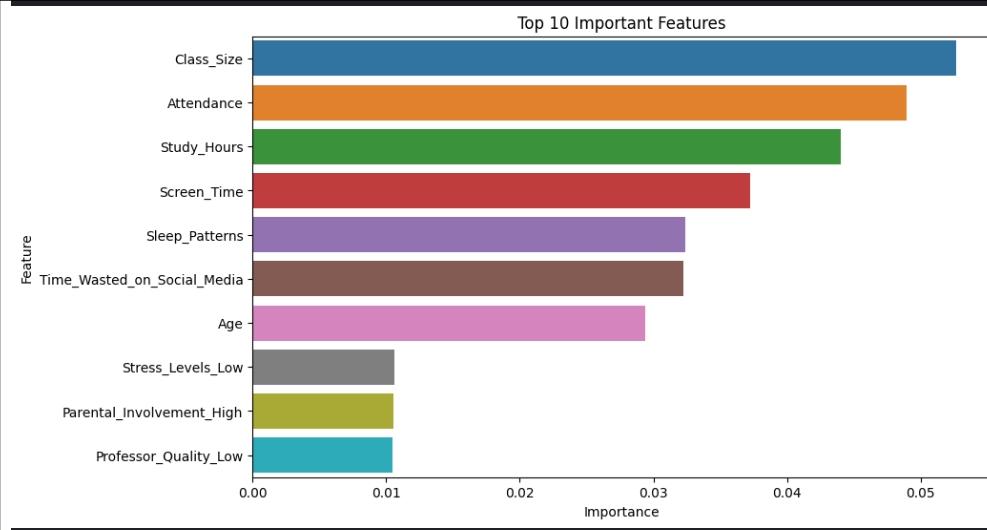
* "**Negative**": Among students in a negative school environment, the most frequent grade is "B," followed by "A," and then "C."
* "**Positive**": The trend is similar for students in a positive school environment, with "B" being the most common, followed by "A," and then "C."
* "**Neutral**": Students in a neutral school environment also follow the same pattern, with "B" being the most frequent, followed by "A," and then "C."

1. **Impact of School Environment on Grades:**

* **Relatively Consistent Patterns:** The distribution of grades across the three school environment categories is relatively consistent. All three groups show a similar pattern of having the highest count of "B" grades, followed by "A" and then "C."
* **Potential Slight Variations:** While the overall patterns are similar, there might be slight variations in the exact proportions of each grade level within each school environment group. However, these differences don't appear to be very large at first glance.

1. **Potential Implications:**

* **School Environment Might Not Be the Dominant Factor:** The data suggests that the school environment alone might not be the most influential factor in determining student grades, as the distributions are relatively similar across all three categories.



1. **Feature Importance Ranking:**

* **Class\_Size** is the most important feature, having the longest bar. This suggests that the size of the class has the strongest influence on the model's predictions (student grades).
* **Attendance** is the second most important feature, indicating a strong correlation between attendance and grades.
* **Study\_Hours** is the third most important feature, which aligns with common expectations about academic success.
* **Screen\_Time** and Sleep\_Patterns also appear to be significant factors.
* **Time\_Wasted\_on\_Social\_Media** has a moderate level of importance.
* **Age** shows a relatively lower importance compared to the top features.
* **Stress\_Levels\_Low**, **Parental\_Involvement\_High**, and **Professor\_Quality\_Low** have the lowest importance among the top 10, but still contribute to the model's predictive power.

1. **Interpretation:**
   * The model suggests that factors related to academic engagement (**Class\_Size**, **Attendance**, **Study\_Hours**) and **lifestyle** (**Screen\_Time**, **Sleep\_Patterns**) are the most influential predictors of student grades.

* While **Age**, **Stress Levels**, **Parental Involvement**, and **Professor Quality** make a contribution, their impact is comparatively smaller.

**Conclusions:**

* **Key Factors Influencing Grades:** The analysis identified several key factors that appear to influence student grades, including **class size**, **attendance**, **study hours**, **screen time**, **sleep patterns**, and **time wasted on social media**. These factors relate to both academic engagement and lifestyle choices.
* **Importance of Holistic Approach:** The findings underscore the importance of a holistic approach to student support. While academic factors like **study habits** and **attendance** are crucial, lifestyle factors like **sleep** and **screen time** also play a significant role. Socio-economic factors, parental education, and school environment, while showing less direct linear influence in some visualizations, are still likely to be contributing factors in the complex system that affects student outcomes.
* **Potential of Targeted Interventions:** The random forest model highlighted the relative importance of different features, suggesting areas where targeted interventions could be most effective.
* **Limitations of Observational Data:** The analysis is based on observational data, which limits the ability to draw causal conclusions. Correlations do not equal causation. Further research is needed to establish causal links between the identified factors and student grades.

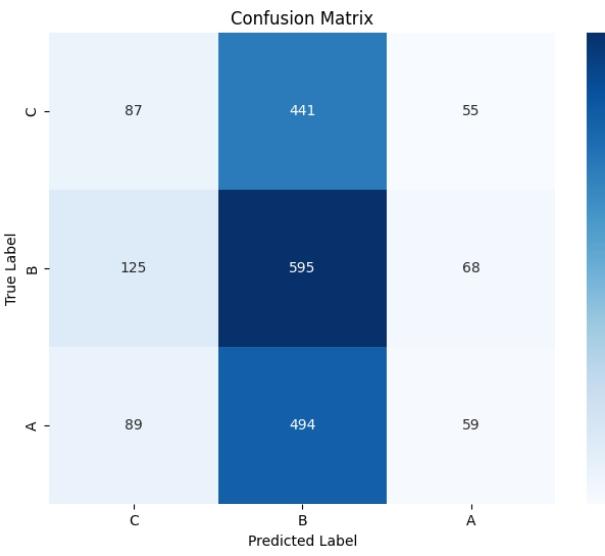
**Recommendations:**

1. **Focus on Academic Engagement and Lifestyle:** Develop programs and initiatives that address both academic engagement and lifestyle factors. This could include:

* **Study skills workshops:** Teach effective study techniques, time management, and note-taking strategies.
* **Attendance improvement programs:** Investigate reasons for poor attendance and implement strategies to encourage regular class attendance.
* **Screen time management workshops:** Educate students about the impact of excessive screen time on academic performance and provide strategies for managing it.
* **Sleep hygiene education:** Raise awareness about the importance of sleep for academic success and provide tips for improving sleep quality.
* **Social media awareness campaigns:** Help students understand the potential distractions of social media and promote responsible usage.

1. **Targeted Support Based on Risk Factors**: Use the identified risk factors (e.g., large class sizes, poor attendance, excessive screen time) to identify students who might be in need of additional support. Offer targeted interventions to these students.
2. **Advocate for Smaller Class Sizes:** Given the strong influence of class size, advocate for smaller class sizes where feasible. Present data and research to school administrators and policymakers to support this recommendation.
3. **Invest in Resources and Mentoring:** Continue investing in resources and mentoring programs, as these can provide valuable support to students. Explore ways to make these programs accessible to all students, regardless of their background.
4. **Further Research to Establish Causality:** Conduct further research to establish causal relationships between the identified factors and student grades. This could involve more rigorous statistical methods or qualitative studies. Consider partnering with academic institutions to conduct this research.
5. **Holistic Support Programs:** Develop holistic support programs that address multiple factors simultaneously. For example, a program might combine study skills workshops with sleep hygiene education and stress management techniques.
6. **Data-Driven Decision Making:** Use data to track the effectiveness of programs and initiatives. Regularly evaluate program outcomes and make adjustments as needed. Continuously collect data on student demographics, academic performance, and lifestyle factors to inform decision-making.
7. **Collaboration with Schools and Families:** Collaborate closely with schools and families to create a supportive learning environment for students. Engage parents in their children's education and work with schools to implement effective interventions.
8. **Long-Term Perspective:** Recognize that improving student outcomes is a long-term process. Be patient and persistent in implementing programs and initiatives. Continuously evaluate and refine strategies based on data and research.

**Limitations**:



1. **Accuracy**:
   * Overall accuracy is calculated as the sum of correctly classified instances divided by the total number of instances.
   * (87 + 595 + 59) / (87 + 441 + 55 + 125 + 595 + 68 + 89 + 494 + 59) = 741 / 2013 ≈ 0.368 or 36.8%
   * This means the model is only accurate about 36.8% of the time, which is not very good.
2. **Precision**:
   * Precision for each class measures how many of the instances predicted as that class were actually correct.
   * Precision for A: 59 / (55 + 68 + 59) = 59 / 182 ≈ 0.324 (32.4%)
   * Precision for B: 595 / (441 + 595 + 494) = 595 / 1530 ≈ 0.389 (38.9%)
   * Precision for C: 87 / (87 + 125 + 89) = 87 / 301 ≈ 0.289 (28.9%)
3. **Recall (Sensitivity or True Positive Rate):**
   * Recall for each class measures how many of the actual instances of that class were correctly predicted.
   * Recall for A: 59 / (55 + 68 + 59) = 59 / 182 ≈ 0.324 (32.4%)
   * Recall for B: 595 / (441 + 595 + 494) = 595 / 1530 ≈ 0.389 (38.9%)
   * Recall for C: 87 / (87 + 125 + 89) = 87 / 301 ≈ 0.289 (28.9%)
4. **F1-Score**:
   * The F1-score is the harmonic mean of precision and recall. It balances both metrics.
   * F1-Score for A: (2 \* 0.324 \* 0.324) / (0.324 + 0.324) ≈ 0.324
   * F1-Score for B: (2 \* 0.389 \* 0.389) / (0.389 + 0.389) ≈ 0.389
   * F1-Score for C: (2 \* 0.289 \* 0.289) / (0.289 + 0.289) ≈ 0.289

**Key Observations and Potential Issues:**

* **Low Accuracy:** The overall accuracy of 36.8% indicates that the model is not performing very well.
* **Confusion between B and C:** The largest confusion is between grades B and C. The model frequently misclassifies students with an actual grade of C as B (441 instances) and vice versa (494 instances). This suggests that the features the model is using might not be effectively distinguishing between these two grades”and we proved that earlier”.
* **Balanced Precision and Recall:** For each class, the precision and recall are the same, leading to identical F1-scores. This indicates a balanced performance in terms of correctly predicting positive cases and finding all actual positive cases, but the overall values are quite low.
* **Possible Data Imbalance:** While not explicitly visible from the confusion matrix alone, if the actual dataset has a significant imbalance in the number of students with different grades (e.g., many more students with B's than A's or C's), this could affect the model's performance.