

Analysis of Factors Influencing Student Academic Performance

Prepared by: Adaobi Cynthia Okonkwo

Tools Used: Excel, Data Visualization Techniques

Category: Educational Data Analytics

1. Outline

This report follows a structured roadmap to present the analysis clearly and logically:

- Introduction
- Story of Data
- Data Splitting and Preprocessing
- Pre-Analysis
- In-Analysis
- Post-Analysis and Insights
- Data Visualizations & Charts
- Recommendations and Observations
- Conclusion
- References & Appendices

2. Introduction

This project examines student academic performance, with a focus on understanding the factors that influence students' final grades. The goal is to identify patterns, uncover key determinants of academic success or poor performance, and provide insights that can support improved educational outcomes and decision-making.

The analysis specifically evaluates academic results across various demographic characteristics, family background, school-related support, and lifestyle factors, including study time, parental education, health status, alcohol consumption, and social activities. The core problem being addressed is the variation in student academic performance and the challenge of identifying which factors most strongly affect students' grades.

A structured student performance dataset was used, containing demographic, social, academic, and behavioral variables. Microsoft Excel served as the primary analytical tool. Key Excel methods applied include data cleaning, pivot tables, charts, sorting, filtering, and descriptive analysis.

3. Story of Data

- **Data Source:** The data originates from a publicly available student academic performance dataset, commonly used for educational research and data analytics projects on Kaggle.com.
- **Data Collection:** The data was collected through a combination of school academic records and structured questionnaires completed by students and their guardians. Academic scores and

attendance records were obtained from school databases, while demographic, family, and lifestyle information were gathered through surveys.

- **Data Structure:** Rows: Each row represents an individual student.

Columns: Variables include school, gender, age, address type, family size, parental status, parental education and occupation, reason for school choice, study time, travel time, past failures, academic support indicators, lifestyle factors, health status, absences, and academic grades (G1, G2, and G3).

- **Important Features and Their Significance:**

- Study Time & Failures – Helps assess the impact of learning habits and academic history on performance.
- Parental Education (Medu & Fedu) – Indicates the influence of family educational background on student achievement.
- School and Family Support – Provides insight into the role of external academic assistance.
- Health, Absences, and Lifestyle Factors – Highlights how well-being and behavior affect learning outcomes.
- G3 (Final Grade) – Serves as the target variable for measuring overall academic performance.

Additional engineered features, such as Average Grade and Grade Improvement, were added to enrich the pattern

- **Limitations:** No limitations were present.

4. Data Splitting and Preprocessing

- **Data Cleaning:** The dataset was reviewed for quality issues, including duplicates, inconsistencies, and incorrect entries. Duplicates were not present, invalid values were not present, and categorical fields were checked to improve reliability.
- **Handling Missing Values:** The dataset contained no missing or null values. Therefore, no imputation or deletion techniques were required, ensuring the completeness and reliability of the analysis.
- **Data Transformations:**
 - Average Grade: A new variable, “Average Grade”, was created by calculating the mean of G1, G2, and G3 for each student. This variable provides an overall view of a student’s academic performance across all grading periods rather than focusing on a single exam score.

- Grade Improvement: A derived variable, “Grade Improvement”, was created by computing the difference between G3 (final grade) and G1 (first period grade). This transformation enables the tracking of students’ academic progress over time. The formula classifies academic progress into three categories:

Improved – when the final grade is higher than the initial grade

Declined – when the final grade is lower than the initial grade

No Change – when both grades are equal

- Categorical Value Recoding: Several variables in the dataset were originally encoded numerically (e.g., 1–5). These values were transformed into their corresponding categorical labels (such as health status, parental education level, family relationships, etc.) to improve readability, interpretation, and accuracy of analysis.
- No normalization or standardization was required, as the analysis was primarily descriptive. However, categorical variables were properly formatted to support sorting, filtering, and pivot table analysis in Microsoft Excel.

- **Data Splitting:** Variables were separated into:

- Independent Variables: School, Sex, Age, Address, Family Size, Parental Status, Mother's Educational Level, Father's Educational Level, Mother's Job, Father's Job, Reason, Guardian, Travel Time, Study Time, Failures, School Support, Family Support, Extra Paid Classes, Activities, Nursery, Higher, Internet, Romantic, Family Relationship, Free time, Go out, alcohol consumption (Dalc and Walc), Health, Absences
- Dependent Variable: G1, G2, Final Grade (G3).

This separation ensures the analysis clearly identifies which demographic, family, school, and lifestyle factors influence students’ academic outcomes.

- **Industry Context:** The dataset represents the education sector, where understanding student performance is essential for improving teaching strategies, academic support systems, and learning outcomes.

- **Stakeholders:** The insights from this analysis are valuable to teachers, school administrators, academic counselors, education policymakers, and educational data analysts, supporting data-driven decision-making within academic institutions.
- **Value to the Industry:** Findings from this project can assist educational institutions in identifying key factors that affect student success, supporting early intervention strategies, improving resource allocation, and enhancing overall academic performance and student retention.

5. Pre-Analysis

The pre-analysis phase provides an initial understanding of the dataset and uncovers early trends before deeper statistical evaluation.

- **Identify Key Trends:** Early exploration of the shipment data revealed several notable trends:
 - Study Time and Performance: Students who spend more time studying generally show higher final grades compared to those with lower study time.
 - Academic History: Students with fewer past academic failures tend to perform better in their final grades.
 - Attendance Patterns: Lower numbers of absences are commonly associated with higher academic achievement.
 - Support Systems: Students receiving school or family support often demonstrate improved academic outcomes.
- **Potential Correlations:** Initial observations suggest possible relationships between variables:
 - Parental Education vs. Academic Performance: Higher levels of parental education (Medu and Fedu) appear to correlate with better student grades.
 - Lifestyle Factors vs. Performance: Increased alcohol consumption (Dalc and Walc) shows a potential negative relationship with academic performance.
 - Study Time vs. Grade Improvement: Higher study time appears to be associated with positive grade improvement over time.
- **Initial Insights:**
 - Early analysis indicates that study habits, attendance, family background, and lifestyle behaviors are likely influential factors in student academic performance.

- Patterns of lower performance are more frequently observed among students with high absences, prior failures, and limited support systems, highlighting areas for deeper investigation.

These insights provide a roadmap for the detailed analysis, allowing focused exploration of the most significant variables affecting student academic outcomes.

6. In-Analysis

This stage represents the core of the analytical process, where deeper exploration of the dataset reveals verified and emerging patterns. Using Excel-based techniques such as PivotTables, correlation checks, conditional formatting, and lookup functions, several insights were uncovered.

- **Unconfirmed Insights (Hypotheses Under Evaluation):** During the analysis, several patterns and relationships were observed that suggest potential influence but require further statistical validation:
 - Study Time vs. Academic Performance: Students with higher study time often achieve better final grades, though the strength of this relationship varies across different age groups and schools.
 - Family and School Support: Students receiving family or school support appear to perform better academically, but this effect may depend on other factors such as parental education and past failures.
 - Lifestyle Behaviors: Higher levels of alcohol consumption (Dalc and Walc) tend to coincide with lower grades, suggesting a possible negative impact on academic outcomes.
 - Attendance and Performance: Increased absences frequently align with reduced academic performance, though a small number of students perform well despite higher absence counts, raising questions for deeper investigation.

These observations form hypotheses that guide further analysis and interpretation.

- **Recommendations (Preliminary and Based on Emerging Patterns):** Although the analysis is ongoing, early patterns allow for initial recommendations:

- Encourage Effective Study Habits: Educational institutions should promote structured study schedules and time management to improve student outcomes.
- Strengthen Support Systems: Schools and families should collaborate to provide academic support, especially for students with prior failures or low performance trends.
- Monitor Attendance and Well-being: Early intervention strategies should be implemented for students with high absences or declining health indicators.
- Promote Healthy Lifestyles: Awareness programs addressing lifestyle behaviors may help reduce negative impacts on academic performance.

These recommendations will be refined once hypotheses are fully validated.

- **Analysis Techniques Used in Excel:** Excel served as the primary tool for exploring and validating relationships in the dataset. Key functionalities included:
 - Pivot Tables to summarize grades across demographic, family, and lifestyle categories.
 - Sorting and Filtering to compare performance across different student groups.
 - Advanced Formulas such as IF, AVERAGE, and calculated fields to create derived variables like *Average Grade* and *Grade Improvement*.
 - Charts and Visualizations to identify trends and patterns in academic performance.

7. Post-Analysis and Insights

This section summarizes the conclusions drawn from the analysis and evaluates whether the initial assumptions identified during the pre-analysis phase were supported by the data. It highlights confirmed relationships, clarifies unexpected outcomes, and consolidates insights gained from the detailed exploration.

- **Key Findings:** Based on the completed analysis, several significant insights emerged:
 - Academic performance is highest and most concentrated among students aged 16-17, while contributions drop sharply at older ages, likely due to fewer students rather than lower ability.
 - Gabriel Pereira has more students overall, and female students outnumber males in both schools, with the gender gap more pronounced at Mousinho da Silveira.

- Most students improved their grades, with over half showing improvement, while fewer students experienced no change or a decline.
- Students with no past failures have the lowest average absences, while those with previous failures, especially combined with poorer health, tend to miss school more, showing that failures and health status are linked to higher absenteeism.
- Most students are from urban areas, with urban students making up the clear majority of the dataset compared to rural students.
- Most students chose their school because of the course offered, with home location and school reputation being secondary reasons.
- Participation in extracurricular activities shows little to no difference in grade improvement, as students with and without activities have nearly identical improvement, no-change, and decline counts.
- Overall, students' average grades steadily improve over time, rising from 11.40 in G1 to 11.57 in G2 and 11.91 in G3.

- **Comparison with Initial Findings**

Comparing the validated results with the early hypotheses reveals several aligned and surprising outcomes:

- Validated Hypotheses: Students with higher study time, consistent attendance, and family or school support generally achieve better grades. Past academic failures negatively impact performance, and higher parental education correlates with improved outcomes.
- Surprises / Counter-Intuitive Results: Some students with higher absences still maintained good grades, suggesting that individual learning habits or external study efforts can offset absenteeism. Participation in extracurricular activities showed little effect on grade improvement, contrary to expectations that involvement would boost performance.
- Insights: Study habits, attendance, family support, and parental education are key determinants of academic success. Monitoring at-risk students, encouraging consistent study practices, and providing targeted support can improve overall academic outcomes. Interventions should consider both academic and non-academic factors to effectively enhance student performance.

8. Data Visualizations & Charts

Visual representations were created in Microsoft Excel to simplify complex data relationships and highlight key patterns related to student performance. A summary of the visualizations and their interpretations is presented below.

- **Visuals**

- **Age Distribution of Average Grade (Histogram)**

A histogram was used to visualize the distribution of average grades across different age groups.

Insight: Students aged 16–17 record the highest average grades, indicating that academic performance peaks during the typical secondary school years. Performance contributions decline at older ages, largely due to fewer students in those age groups rather than a sharp drop in individual ability.

- **Gender Population by School (Clustered Column Chart)**

This chart shows the comparison of male and female student populations in Gabriel Pereira and Mousinho da Silveira schools.

Insight: Gabriel Pereira School has a higher total student population and therefore records higher counts for both male and female students, making it the primary contributor to overall enrollment, while Mousinho da Silveira has a smaller population with fewer students across both genders.

- **Impact of Parental Status on Grade Improvement (Clustered Column Chart)**

This visualization shows how students' grade improvement varies by parental status, highlighting differences in academic outcomes across the groups.

Insight: Students with parents living together show the highest level of grade improvement, while those from other parental status categories record lower improvement levels, indicating that parental status has a noticeable influence on academic performance.

- **Impact of Absences and Health on Student Failures (Stacked Bar Chart)**

This visualization shows how student failure rates vary with different levels of absences across health categories, highlighting the combined impact of attendance and health status on academic performance.

Insight: Higher absences are strongly linked to academic struggles, particularly among students with poor health. Students with no failures have the lowest absences, while those with multiple failures and poor health show the highest, identifying a key at-risk group needing both academic and wellness support.

- **Student Distribution by Location (Donut Chart)**

This visualization shows the proportion of students living in rural and urban areas, indicating that the majority of students come from urban locations.

Insight: Urban students comprise the majority (~70%), indicating higher enrollment or data representation in urban areas, which may reflect geographical disparities in educational access or outcomes.

- **Distribution of Students by School Selection Reason (Pie Chart)**

This visualization shows the distribution of students based on their reason for choosing a school (Course, Home, Reputation, Other).

Insight: Most students selected their school for the course offered, while home location and school reputation are secondary factors, highlighting the primary role of academic programs in school choice.

- **Grade Improvement Outcomes by Participation in Activities (Stacked Column Chart)**

This visualization shows students' grade improvement outcomes (Improved, No Change, Declined) based on whether they participate in extracurricular activities.

Insight: Participation in activities has little effect on grade improvement, as students with and without activities show similar patterns of improvement, no change, and decline.

- **Progression of Grade Performance (Area Chart)**

This visualization shows the progression of students' average grades across grading periods (G1 → G2 → G3).

Insight: Students' grades steadily improve over time, indicating gradual academic progress throughout the school year.

- **Students' Academic Performance Analysis (Excel Dashboard)**

A consolidated dashboard was developed using slicers, pivot charts, and KPI indicators to allow interactive exploration of: Age Distribution of Average Grade, Gender Population by School, Impact of Parental Status on Grade Improvement, Student Distribution by Location, Distribution of Students by School Selection Reason, Grade Improvement Outcomes by Participation in Activities, Progression of Grade Performance.

Dashboard Insights: The dashboard provides an at-a-glance view of how student, family, and school variables interact with academic performance, helping stakeholders quickly identify at-risk students and factors contributing to lower grades.

9. Recommendations and Observations

This section translates the findings from the student performance analysis into actionable recommendations that can help improve academic outcomes, student engagement, and overall school performance.

○ Actionable Insights

- Promote structured study schedules and time management to help students improve grades, particularly for those showing minimal grade improvement.
- Provide targeted interventions for students with multiple past failures, high absences, or poor health, as they are more likely to struggle academically.
- Collaborate with families and teachers to offer consistent guidance and academic assistance, especially for students from less supportive backgrounds.
- Implement attendance monitoring and early warning systems to address absenteeism before it impacts performance.

Optimizations or Business Decisions

- Focus academic support programs and counseling services on the age groups and students showing the highest risk (e.g., multiple failures, high absences).
- Adjust teaching methods and study materials to account for varying student backgrounds, parental education, and learning needs.

- Use student performance dashboards to continuously track improvements, enabling proactive measures rather than reactive support.

Unexpected Outcomes

- Participation in activities had little effect on grade improvement, suggesting that time spent in activities should be balanced with study commitments.
- Some students with good or very good health still had moderate absences, indicating that engagement or motivation factors may influence attendance independent of health.

10. Conclusion

This section summarizes the key findings, acknowledges limitations, and outlines directions for future analysis, providing closure to the report while highlighting actionable insights.

○ **Key Learnings**

- Students aged 16–17 contribute most to overall academic performance, with grades improving steadily over time.
- Study time, attendance, family and school support, and parental education are key factors influencing academic success.
- Higher absences and past academic failures, particularly among students with poor health, are strongly linked to lower performance, identifying at-risk groups.
- Participation in extracurricular activities showed minimal impact on grade improvement, while parental status influences outcomes, with students from parents living together performing better.

○ **Limitations**

- Some age groups have very few students, which may skew insights for older ages.
- Data on student motivation, teaching quality, or socio-economic background was not included, limiting understanding of all factors affecting performance.
- The analysis is primarily descriptive and does not include predictive modeling or causal inference.

○ Future Research

- Collect additional data on socio-economic factors, teacher engagement, and learning resources to better understand their impact on performance.
- Explore predictive models to identify students at risk of failing or declining grades early.
- Investigate interactions between extracurricular activities, motivation, and academic performance to refine support strategies.
- Analyze longitudinal data across multiple academic years to track trends and improvements over time.

11. References

This section provides supporting materials and resources for readers who want to explore the analysis in more detail. It ensures transparency, reproducibility, and gives credit to sources and tools used.

- Dataset Source: [Student Performance Dataset](#)
- Tools Used: Microsoft Excel (Pivot Tables, Charts, Dashboard).
- External Resources:
 - Additional literature on educational performance analysis and statistical methods.
 - Online tutorials and documentation for Excel analytical functions.

○ Appendices

- Pivot Tables: Detailed tables showing distributions of students by age, gender, school, parental status, health, and absences.
- Charts and Visualizations: Histogram of age vs. average grade, pie chart of school selection reasons, stacked columns for grade improvement by activity participation, and clustered bar charts for absences vs. health.
- Formulas and Functions: Step-by-step explanations of key Excel formulas used, including:
 $\text{IF } () \text{ to append Grade Improvement } (=IF([@G3]-[@G1]>0,"Improved", IF([@G3]-[@G1]<0,"Declined","No Change")))$ and Average Grade: $(=AVERAGE([@G1], [@G2], [@G3]))$.

Pivot Table setups for aggregating average grades, absences, failures, and categorical distributions to support analysis conclusions.

Dashboard Screenshots: Illustrations of the consolidated dashboard displaying KPIs and critical metrics for stakeholders.