Introduction

This report presents a comprehensive data analysis of student performance factors with a focus on understanding and predicting Grade Point Average (GPA). Using the provided dataset Student_performance_data_new.csv (2,392 students), we explore how various features – such as demographics, study habits, attendance, parental factors, and extracurricular involvement – relate to student GPA. We perform data cleaning and preprocessing, descriptive statistics, visual exploratory analysis, correlation and feature importance assessments, and predictive modeling (linear regression), culminating in key findings and recommendations. The goal is to identify the most important variables influencing GPA, supported by evidence and reasoning, and to build a model that can estimate a student's GPA from the available features.

Objective of the Project

The primary goal of this data analytics project is to uncover key insights into the factors influencing student performance, with a specific focus on GPA (Grade Point Average) as the target variable. By analyzing behavioral, academic, and demographic data, we aim to identify patterns that can help improve academic outcomes.

Stakeholder Focus: Academic Advisors & Curriculum Committee

For this project, we've chosen to focus on Academic Advisors and the Departmental Curriculum Committee as our primary stakeholders. These individuals are directly involved in guiding students, shaping academic policies, and designing interventions to support student success.

Why This Stakeholder?
Academic Advisors and Curriculum Committees play a critical role in student development. They are responsible for:

- Monitoring student progress
- Recommending academic support
- Designing workshops and training sessions
- Creating data-driven academic policies

Our insights will equip them with evidence-based understanding of the underlying factors that impact GPA, helping them take more informed and targeted actions.

About Dataset

The Student Performance Data contains information about 2,392 students and various factors that may influence their academic success. The dataset includes demographics, study habits, parental influence, and participation in extracurricular activities. It focuses on analyzing factors contributing tostudents' Grade Point Average (GPA), which is the primary target variable.

Dataset Overview:

Data Name: Student Performance Data
Industry / Domain: Education
Author Name/Type: Rabie El Kharoua (Independent Researcher / Data Enthusiast)
Creation Date: January 3, 2024
Duration: Academic Year 2022–2023

Student Dataset Description

Variable	Type	Description	Range / Values	Notes / Statistics
StudentID	Identifier	A unique identifier assigned to each student	N/A	Not used in analysis
Age	Continuous	Age of the student	15 to 18 years	
Gender	Binary	Gender of the student	0 = Male, 1 = Female	51% Female, 49% Male
Ethnicity	Categorical	Ethnic background (coded)	0: Caucasian 1:(African American) 2: Asian 3: Other	0 (~50%), 1 (~21%), 2 (~20%), 3 (~9%)
ParentalEducation	Ordinal	Highest parental education level	0: None 1: High School 2: Some College 3: Bachelor’s 4: Higher	Median = 2 → Many parents have mid-level education
StudyTimeWeekly	Continuous	Hours per week spent studying	~0 to 20 hours	Mean ≈ 9.8 hours/week
Absences	Integer	Number of school absences	0 to 29 days	Mean ≈ 14.5, Median = 15
Tutoring	Binary	Whether the student receives tutoring	0 = No, 1 = Yes	About 30% receive tutoring
ParentalSupport	Ordinal	Level of parental support	0: None 1: Low 2: Moderate	Average ≈ 2.12

			3: High 4: Very High	(moderate support)
Extracurricular	Binary	Participation in extracurricular activities	0 = No, 1 = Yes	~71.6% participate
GPA	Continuous	Grade point average (target variable)	scale from 0.0 to 4.0	Mean ≈ 1.906, Std Dev ≈ 0.915
GradeClass	Categorical	Performance category (likely ranked group)	0: 'A' (GPA >= 3.5) 1: 'B' (3.0 <= GPA < 3.5) 2: 'C' (2.5 <= GPA < 3.0) 3: 'D' (2.0 <= GPA < 2.5) 4: 'F' (GPA < 2.0)	Strong inverse correlation with GPA; used for descriptive, not predictive modeling

```
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   StudentID               2392 non-null   int64
 1   Age                     2392 non-null   int64
 2   Gender                  2392 non-null   int64
 3   Ethnicity               2392 non-null   int64
 4   ParentalEducation       2392 non-null   int64
 5   StudyTimeWeekly         2392 non-null   float64
 6   Absences                2392 non-null   int64
 7   Tutoring                2392 non-null   int64
 8   ParentalSupport         2392 non-null   int64
 9   Extracurricular_Updated 2392 non-null   int64
 10  GPA                     2392 non-null   float64
 11  GradeClass              2392 non-null   int64
dtypes: float64(2), int64(10)
```

```
Ethnicity
0    1207
1     493
2     470
3     222
Name: count, dtype: int64
```

```
ParentalEducation
2    934
1    728
3    367
0    243
4    120
Name: count, dtype: int64
```

Feature	Count	Mean	Std	Min	25%	50% (Median)	75%	Max
Age	2392	16.47	1.12	15.0	15.0	16.0	17.0	18.0
Gender	2392	0.51	0.5	0.0	0.0	1.0	1.0	1.0
Ethnicity	2392	0.88	1.03	0.0	0.0	0.0	2.0	3.0
ParentalEducation	2392	1.75	1.0	0.0	1.0	2.0	2.0	4.0
StudyTimeWeekly	2392	9.77	5.65	0.0	5.04	9.71	14.41	19.98
Absences	2392	14.54	8.47	0.0	7.0	15.0	22.0	29.0
Tutoring	2392	0.3	0.46	0.0	0.0	0.0	1.0	1.0
ParentalSupport	2392	2.12	1.12	0.0	1.0	2.0	3.0	4.0
Extracurricular_Updated	2392	0.72	0.45	0.0	0.0	1.0	1.0	1.0
GPA	2392	1.91	0.92	0.0	1.17	1.89	2.62	4.0
GradeClass	2392	2.98	1.23	0.0	2.0	4.0	4.0	4.0

METHODOLOGY

1. Problem Identification

Identify key academic challenges by analyzing factors that influence students' Grade Point Average (GPA). The main objective is to derive meaningful insights to support strategies for improving academic performance.

2. **Data Understanding and Requirement Gathering**

Study the structure and content of the provided dataset, which includes 12 attributes of 2,392 students, focusing on:

- Student demographics
- Study habits and attendance
- Parental background and support
- Academic outcomes (GPA and GradeClass)

3. **Data Cleaning and Preprocessing**

- Handle missing or inconsistent values
- Encode categorical and ordinal data
- Normalize or scale numerical variables (like StudyTimeWeekly, Absences)
- Used Power Query to merge sports, music, volunteering as Extracurricular activities.

4. **Exploratory Data Analysis (EDA)**

Perform descriptive statistics and visual exploration to:

- Understand distribution and patterns
- Identify outliers or skewed data
- Examine relationships between GPA and factors like study time, absences, parental support, and extracurriculars

5. **Data Visualization**

Create intuitive visualizations to support insights:

- Histograms of GPA
- Box plots for comparing GPA by gender or tutoring
- Heatmaps for correlation

6. **Insights and Interpretation**

Interpret analytical results to discover:

- Which factors have the strongest impact on GPA

- How behaviors like study time or tutoring help performance
- Whether parental support compensates for low parental education

7. **Recommendations to Stakeholders**

Deliver actionable suggestions based on findings:

- Educators: Monitor students with low study time or high absences
- Parents: Encourage consistent involvement
- Administrators: Expand access to tutoring and extracurriculars for lower-performing groups

8. **Report Compilation**

- Document the complete process including:
- Dataset overview and preprocessing
- Key visualizations and EDA summaries
- Insights and takeaways
- Final suggestions for intervention strategies

Data Preprocessing

- **Dataset completeness**
  No missing values in any column; therefore, no imputation was required.
- **Categorical encoding**
  Binary variables are already coded as 0/1 and multi-category variables as integer codes. No further encoding was needed, though one-hot encoding of purely nominal categories (e.g., Ethnicity) could be applied in a more rigorous model.
- **Outlier check**
  All numeric features lie within their expected ranges (GPA 0–4, StudyTimeWeekly 0–20, Absences 0–29). A few very low GPA values correspond to extremely high absences, but no values were extreme enough to warrant removal.
- **Removed columns (using power query)**
  The **Sports, Volunteering and Music** column were embedded in **Extracurricular_Updated** column with **Extracurricular** column and dropped at import as they were redundant.

```arduino
if [Extracurricular] = 0 and
   ([Sports] = 1 or [Music] = 1 or [Volunteering] = 1)
then 1
else [Extracurricular]
```

Before

StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular	Sports	Music	Volunteering	GPA	GradeClass
1001	17	1	0	2	19.83372281	7	1	2	0	0	1	0	2.929196	2
1002	18	0	0	1	15.40875606	0	0	1	0	0	0	0	3.042915	1
1003	15	0	2	3	4.210569769	26	0	2	0	0	0	0	0.112602	4
1004	17	1	0	3	10.02882947	14	0	3	1	0	0	0	2.054218	3
1005	17	1	0	2	4.672495273	17	1	3	0	0	0	0	1.288061	4
1006	18	0	0	1	8.191218545	0	0	1	1	0	0	0	3.084184	1
1007	15	0	1	1	15.60168047	10	0	3	0	1	0	0	2.748237	2
1008	15	1	1	4	15.42449631	22	1	1	1	0	0	0	1.360143	4
1009	17	0	0	0	4.562007558	1	0	2	0	1	0	1	2.896819	2
1010	16	1	0	1	18.44446636	0	0	3	1	0	0	0	3.573474	0
1011	17	0	0	1	11.85136366	11	0	1	0	0	0	0	2.147172	3
1012	17	0	0	1	7.598485819	15	0	2	0	0	0	1	1.559595	4
1013	17	0	1	1	10.03871162	21	0	3	1	0	0	0	1.520078	4
1014	17	0	1	2	12.10142507	21	0	4	0	1	0	0	1.751581	4

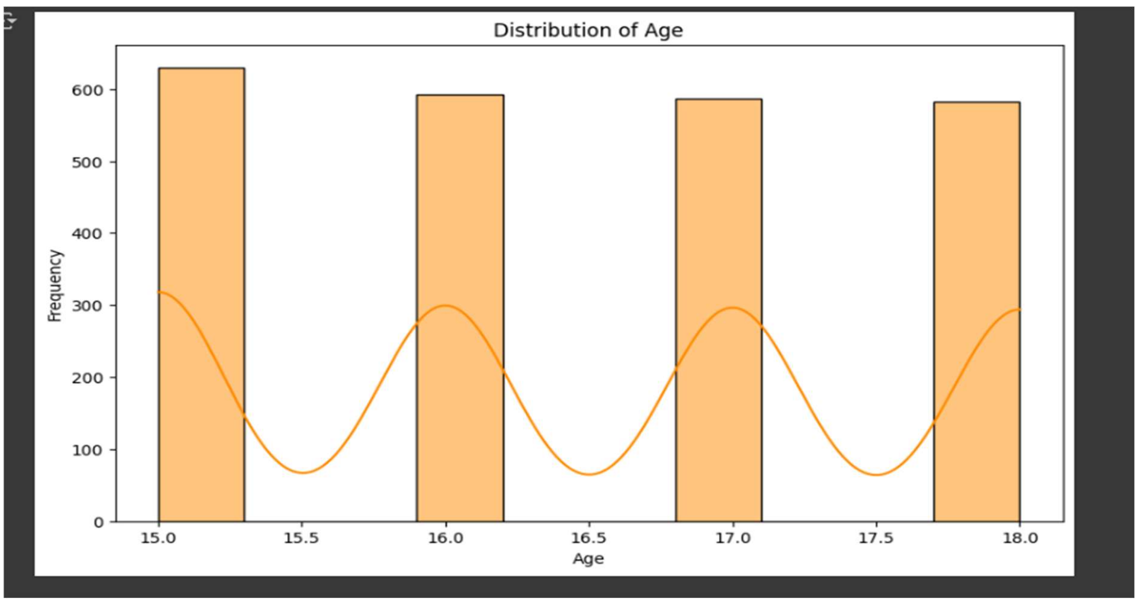
After

StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular_Updated	GPA	GradeClass
1001	17	1	0	2	19.83372281	7	1	2	1	2.929195592	2
1002	18	0	0	1	15.40875606	0	0	1	0	3.042914833	1
1003	15	0	2	3	4.210569769	26	0	2	0	0.112602254	4
1004	17	1	0	3	10.02882947	14	0	3	1	2.05421814	3
1005	17	1	0	2	4.672495273	17	1	3	0	1.288061182	4
1006	18	0	0	1	8.191218545	0	0	1	1	3.084183614	1
1007	15	0	1	1	15.60168047	10	0	3	1	2.748237415	2
1008	15	1	1	4	15.42449631	22	1	1	1	1.360142712	4
1009	17	0	0	0	4.562007558	1	0	2	1	2.89681919	2
1010	16	1	0	1	18.44446636	0	0	3	1	3.57347421	0
1011	17	0	0	1	11.85136366	11	0	1	0	2.147171625	3
1012	17	0	0	1	7.598485819	15	0	2	1	1.559594519	4
1013	17	0	1	1	10.03871162	21	0	3	1	1.520077815	4

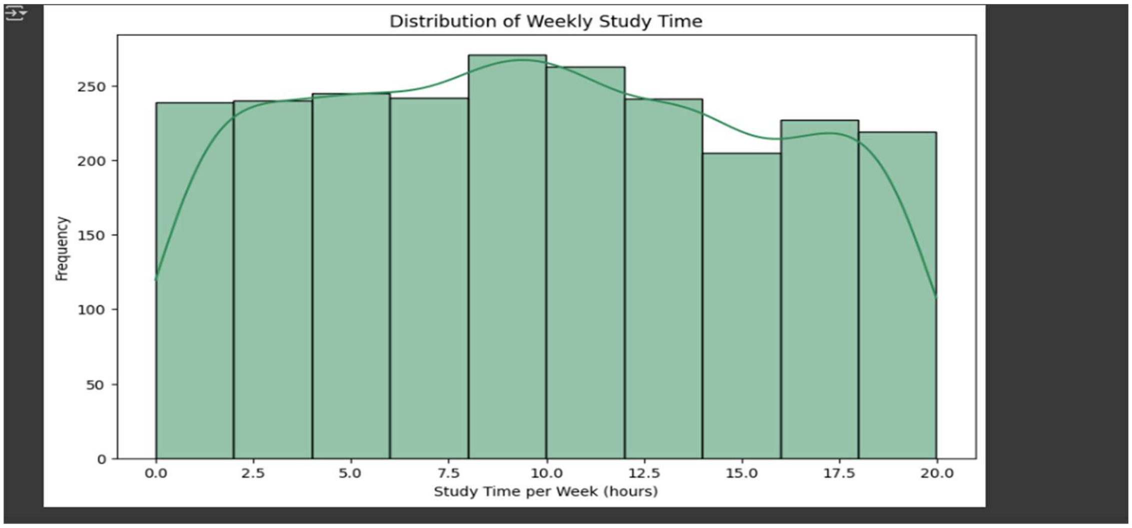
Exploratory Data Analysis (EDA)
&
Data Visualization

1. Distribution of Continuous Variables

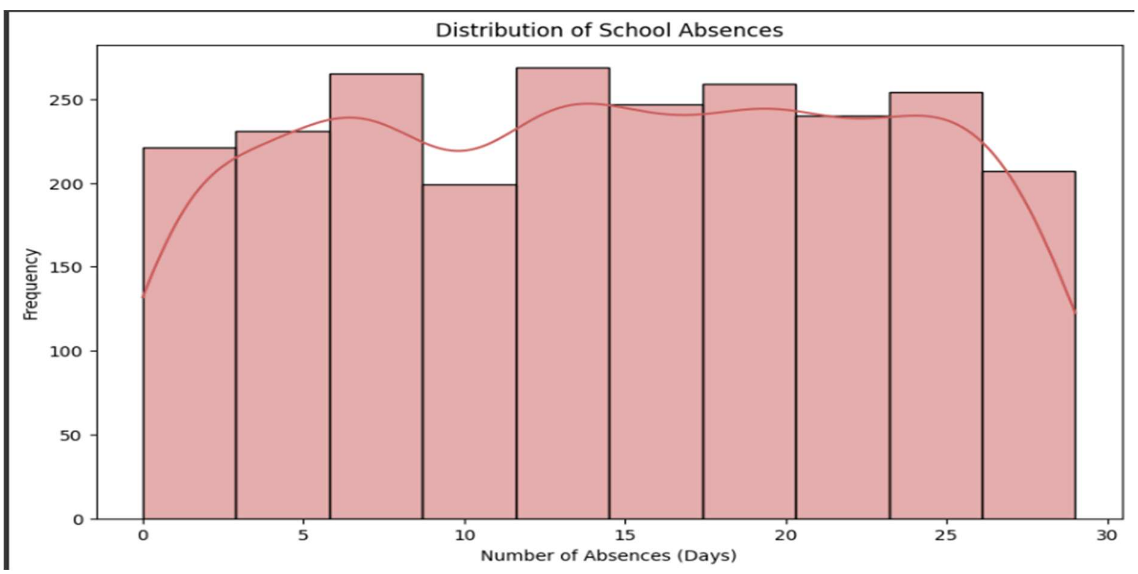
We visualize the distributions of **Age**, **Study Time per Week**, **Absences**, and **GPA**
From the histograms, we observe:
1.1 Age: Each age 15, 16, 17, 18 has similar counts (around 600 each), so no age group dominates the sample.



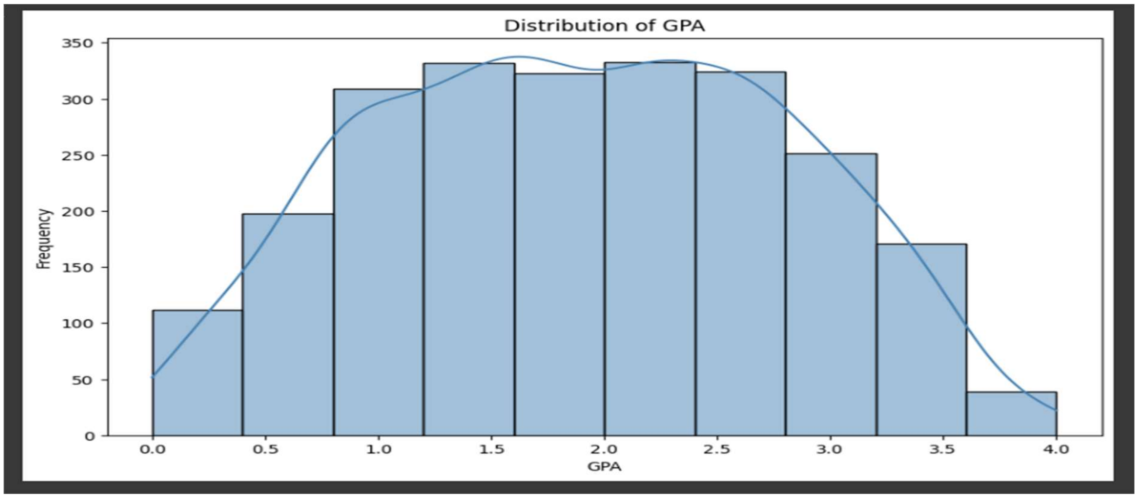
1.2 Study Time: There is no strong skew – students are about as likely to study very little as they are to study a lot. A slight bump around 10–15 hours suggests many aim for roughly 2 hours per weekday. There is a small proportion of extremely low study hours (~0–1 hr) and some who study nearly 20 hours.



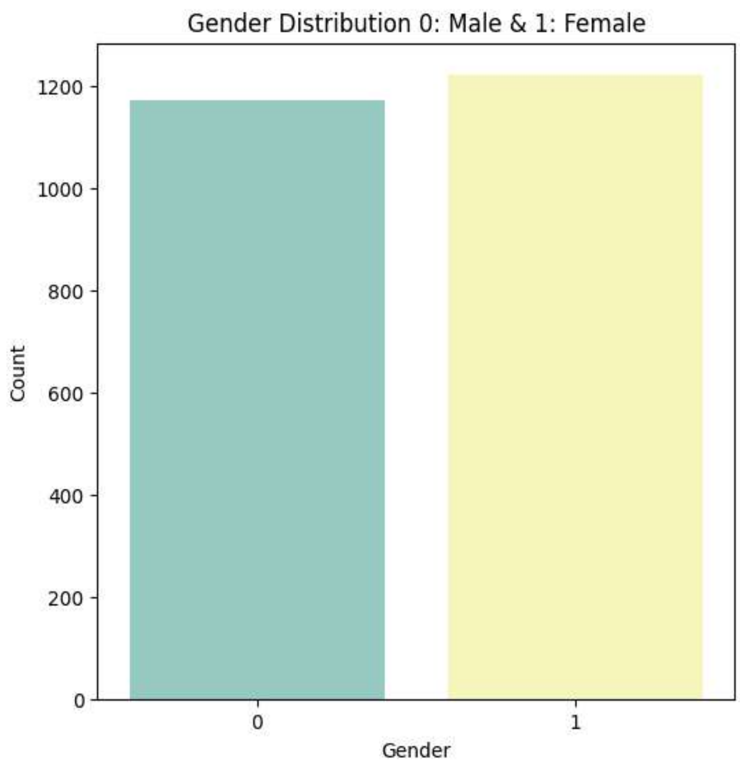
1.3 Absences: The nearly uniform distribution from 0 to 30 absences is noteworthy. It means there's a huge variation in attendance: some students have perfect attendance (0 days missed), whereas some missed about 4 weeks of school (30 days). Many students are in between, with the median at 15 days absent. This likely has a direct impact on learning – missing class can **directly affect students' academic performance**. We expect to see GPA declining as absences increase, a trend widely observed in educational research (e.g., college students with no absences had significantly higher GPAs than those with many absences).



1.4 GPA: The GPA histogram confirms that student performance varies widely. While the distribution peaks around 2.0 (C average), a non-negligible number of students have very low GPAs (<1.0, failing). There are also high achievers (GPAs 3.5–4.0) but they are fewer. The spread (σ ~0.92) indicates substantial performance differences across the student body. This variance in GPA is what we aim to explain and predict using the other features.



1.5 Gender Distribution - The gender countplot shows a nearly equal number of male (0) and female (1) students, ensuring balanced representation. This minimizes bias in analysis and allows for reliable gender-based comparisons in areas like GPA, absences, and study time. As a well-distributed categorical feature, gender can be meaningfully explored in relation to other academic factors.

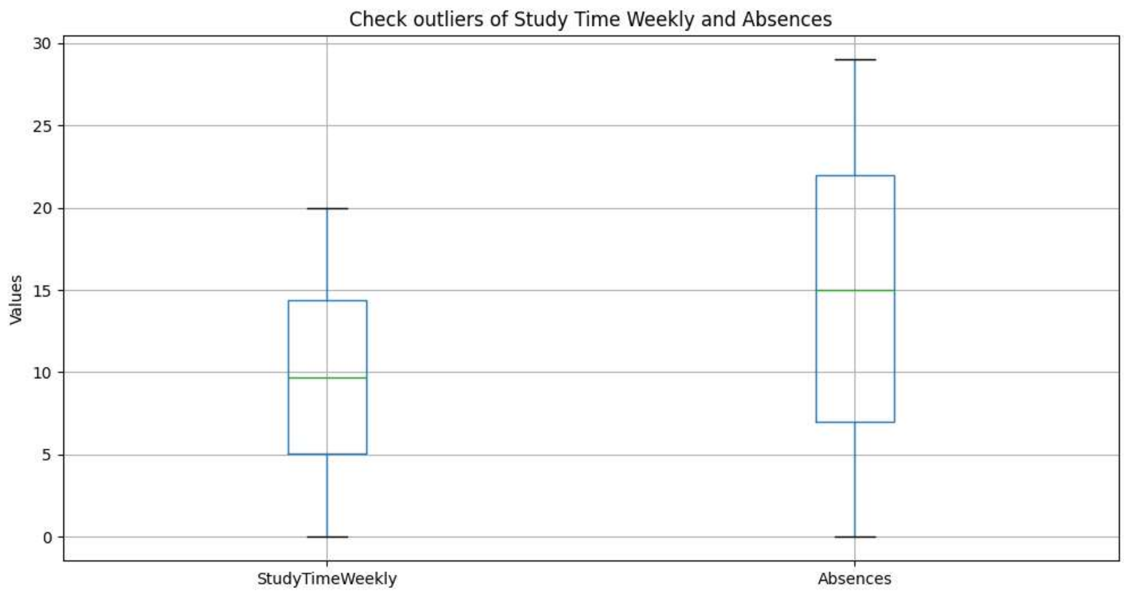


2.1 How do variations in study time and school absences influence academic performance (GPA)?

General Trend Observed:

- **StudyTimeWeekly**:
  - The median is somewhere around 9–10 hours/week.
  - Several **outliers on the higher end** (students studying more than ~18–20 hours).
  - These could represent **very high-performing** or **pressured** students.

- **Absences**:
  - Median around 14–15 absences.
  - **Significant number of outliers** on the upper end — some students have **very high absence rates** (25+ days).
  - Indicates potential academic risk or disengagement.



Conclusion:

- The dataset contains **noticeable outliers** in both study time and absences.
- These data points should be handled thoughtfully:
  - You might **retain** them if they represent real behaviors.
  - Or you could **filter or adjust** them during preprocessing depending on your modeling goals.

2.2 How does study time weekly affect GPA?

General Trend Observed:

- Students studying 0–5 hours/week have the lowest average GPA.

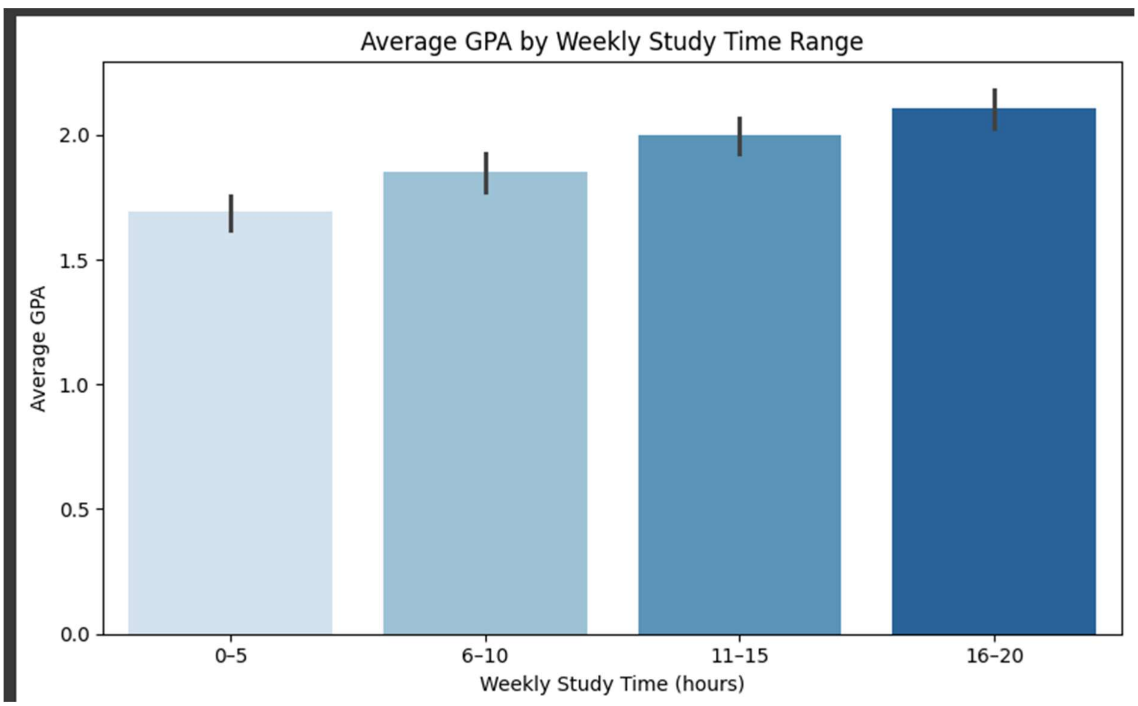
- GPA improves steadily as weekly study time increases, peaking in the 11–15 hour range.
- Students studying 16–20 hours/week may still maintain a high GPA, but the improvement tends to level off, suggesting a plateau effect.

Key Takeaways:

- 6–10 and 11–15 hours/week appear to be the most effective ranges for study time, balancing academic performance without likely burnout.
- Studying less than 5 hours is associated with significantly lower academic performance.
- Excessive studying (16–20 hours) does not necessarily guarantee a higher GPA, possibly due to diminishing returns or other factors like fatigue or inefficient study methods.

Conclusion:

- Moderate study time (around 10–15 hours/week) is associated with optimal GPA outcomes. Simply increasing hours isn't always better — study quality, rest, and balance also matter.



2.3 Is there a significant difference in GPA based on gender?

General Trend Observed:

- The boxplot displays the distribution of GPA for two gender categories:
    - **0** = Female
    - **1** = Male
- Typically, you'll see the **median GPA (the line inside each box)** and how the GPA values are spread out for each gender.

- If the **female group (0)** shows a **higher median** or a **more compact spread near higher GPAs**, it suggests better performance among females.

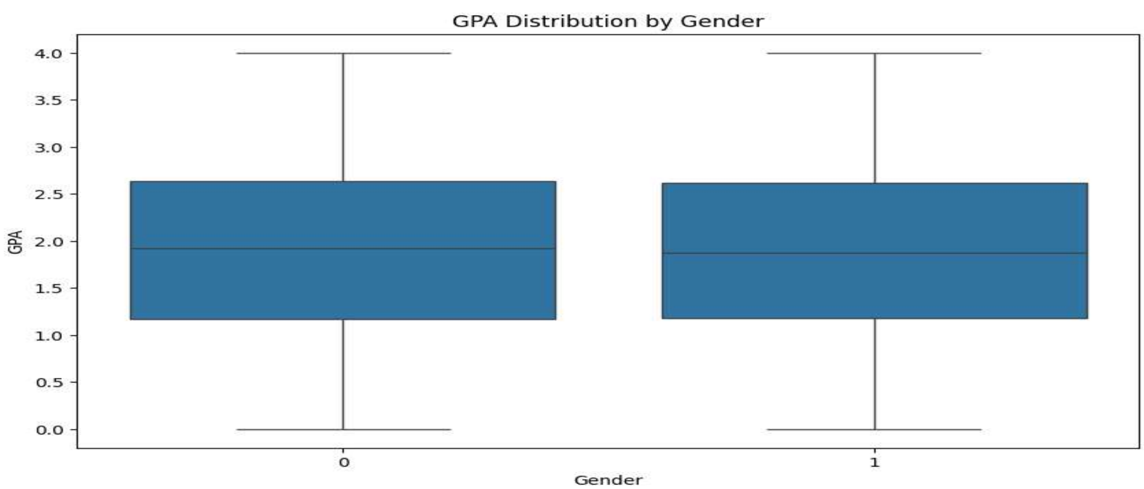
Key Takeaways:

- **Females (Gender = 0)** may have:
  - Higher or more consistent GPA scores.
  - A tighter interquartile range (less variability).

- **Males (Gender = 1)** may show:
  - Slightly lower or more varied GPA scores.
  - More outliers (extreme values), if present.

Conclusion:

- There may be a **significant difference in GPA based on gender**, with females tending to perform slightly better.
- This difference could be further supported by running a **t-test** to assess statistical significance
- Understanding these trends helps tailor academic support strategies based on gender.



2.4 How do extracurricular activities impact GPA?

- The correlation coefficient between Extracurricular participation and GPA is 0.11.

Interpretation:

- This is a positive, but weak correlation.
- It suggests that students who participate in extracurricular activities (like sports, music, volunteering) tend to have slightly higher GPAs, on average.
- However, the relationship is not strong — meaning extracurriculars alone don't guarantee better academic performance.

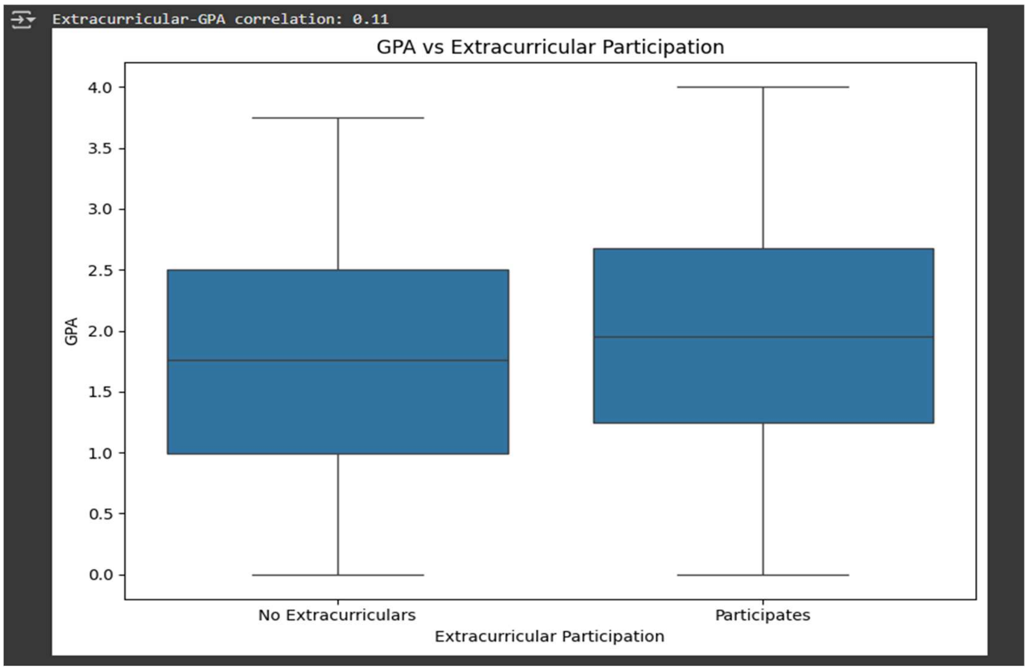
Students engaged in extracurriculars may benefit from:

- Better time management

- Improved discipline and motivation
- A more balanced lifestyle, reducing stress and burnout

Final Insight:

Participation in extracurricular activities shows a **slight academic benefit**, but GPA is shaped by a **complex mix of factors**. Encouraging a **well-rounded lifestyle** may support both academic and personal development.



2.5 How does parental education level correlate with student GPA?

Ans: Since ParentalEducation is ordinal (e.g., 0 = lowest, 4 = highest), and GPA is continuous, Spearman's rank correlation is most appropriate. It captures monotonic relationships (not just linear) and handles ranked/ordered data well.

Spearman correlation between Parental Education and GPA: -0.03 (p-value = 0.0902)

Spearman Correlation = -0.03
Direction: Slightly negative → suggests a tiny inverse relationship between parental education and GPA.
Magnitude: −0.03 is very close to zero → essentially no monotonic relationship.

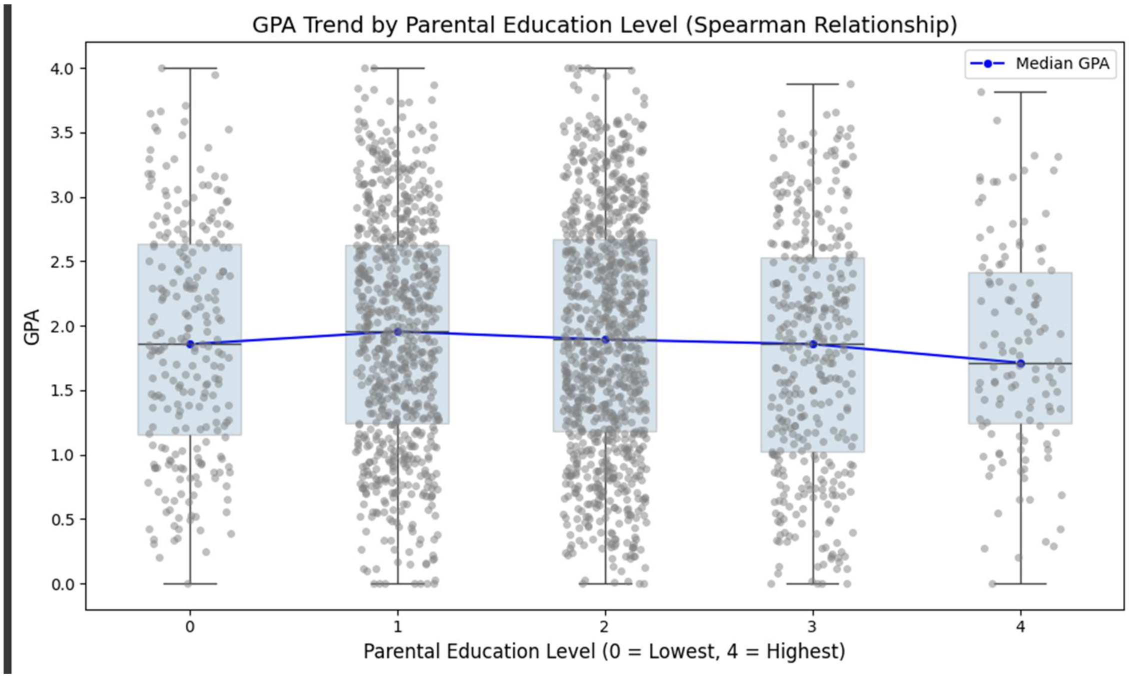
Spearman handles ranked/ordinal data well, but here it's showing no meaningful trend.

p-value = 0.0902
This is above the common significance threshold (0.05), meaning:
The correlation is not statistically significant.

You can’t confidently say there's a real relationship in the population — the result could easily be due to random chance.

Conclusions:-

- Although we might expect higher parental education to boost student GPA, this dataset doesn’t support that assumption.
- The correlation is near zero, and not statistically significant → no reliable relationship found.



2.6 Does receiving tutoring correlate with higher GPAs?

Correlation coefficient = 0.15

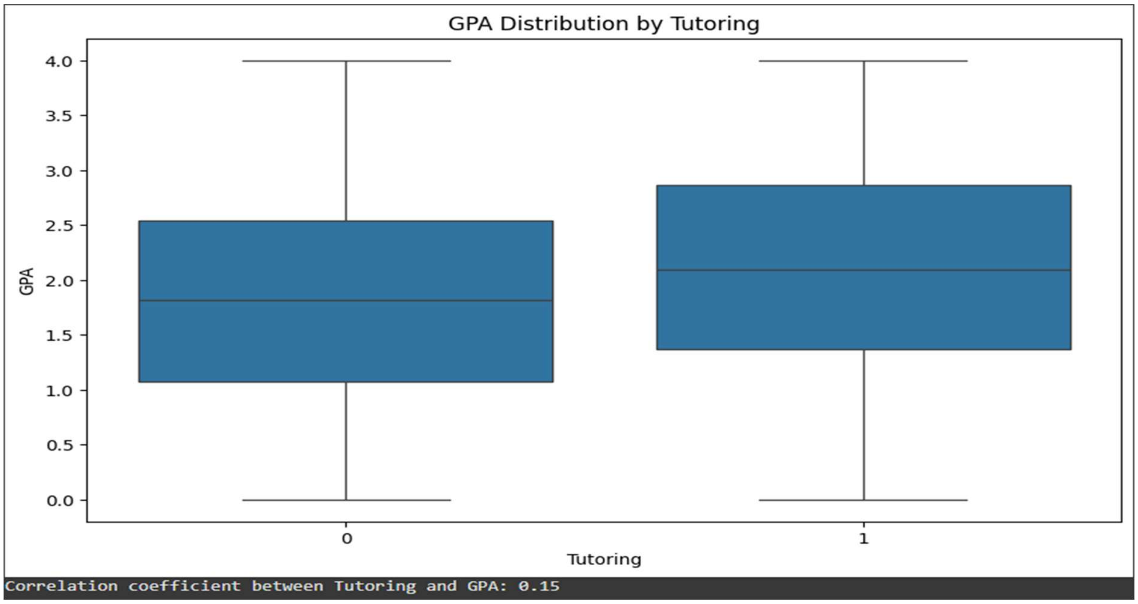
Direction: Positive — suggests that students who receive tutoring tend to have slightly higher GPAs.

Magnitude: 0.15 is a small (but positive) correlation — not strong, but not meaningless either.

Tutoring is somewhat associated with better academic performance, but not a strong predictor on its own.

Conclusions:

- Tutoring provides additional academic support → helps GPA.
- Students struggling more often seek tutoring → could also reflect efforts to improve poor performance.
- A correlation of 0.15 suggests that tutoring may have a mild positive effect on GPA, but it's likely part of a larger picture with multiple influencing factors.



2.7 How does the number of school absences influence students' GPA?

Ans:-

The **Spearman correlation of -0.93** between **Absences** and **GPA** suggests a **very strong negative monotonic relationship**.

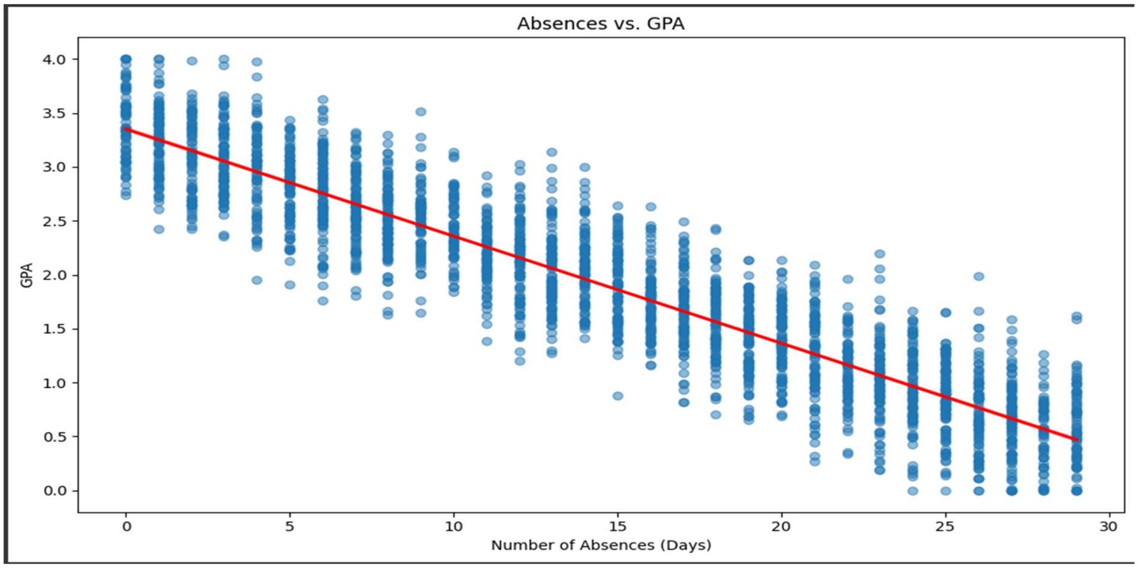
- **Why Spearman?** It captures **monotonic** relationships, even if not linear.
- If the value is negative (e.g., **-0.3 to -0.5**), it indicates a **moderate negative correlation**: more absences → lower GPA.

Interpretation:

- **Strong negative correlation**: As the number of **absences increases**, the **GPA tends to decrease** significantly.
- This indicates that **students who miss more school tend to have much lower GPAs**, with a near **perfect negative monotonic relationship**.
- The **Spearman correlation** is robust to non-linear relationships, and with a value as low as **-0.93**, it indicates that the pattern is not just coincidental but **consistent** across the dataset.

Conclusion:

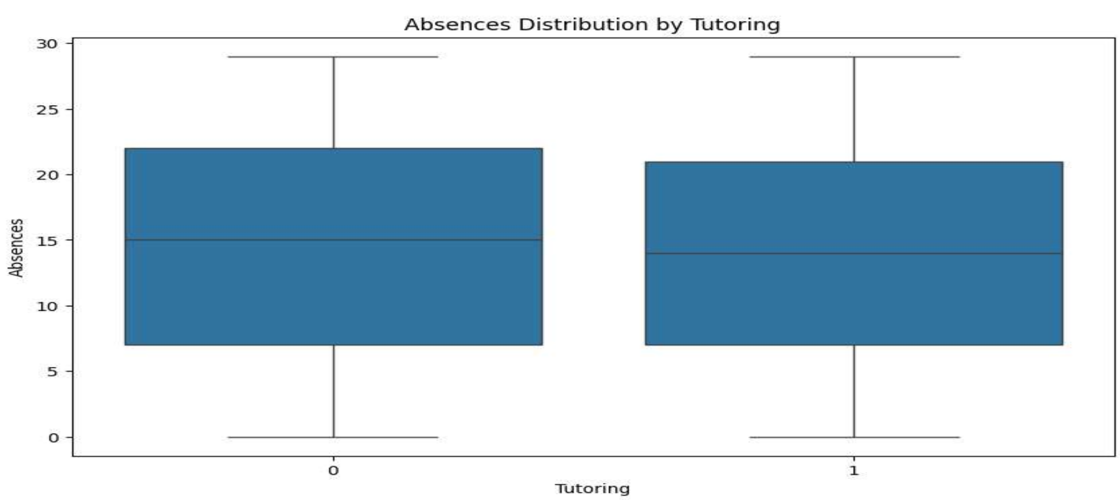
Absenteeism has a strong negative impact on GPA, suggesting that consistent school attendance is crucial for academic success. Students who miss school are likely to miss out on important learning opportunities, which directly affects their grades.



2.8 What is the relationship between tutoring and absences?

The boxplot comparing absences between students with (`Tutoring = 1`) and without (`Tutoring = 0`) tutoring reveals a modest but visible trend: students receiving tutoring generally have **slightly fewer absences**, as shown by a lower median and a tighter interquartile range.
This suggests that students with tutoring may be more engaged or receive additional support that encourages consistent attendance.

Pearson correlation is appropriate here because it measures the **linear relationship** between a **binary variable** (Tutoring: 0 or 1) and a **continuous variable** (Absences). This is a valid and commonly used method for assessing such associations



Conclusion:

Correlation coefficient between Tutoring and Absences: -0.02

This visual impression is backed by the **correlation coefficient of -0.02**, indicating an **extremely weak (almost negligible) negative relationship** between tutoring and absences.

2.9 How does GPA vary across different ethnic groups?

GPA appears relatively consistent across different ethnic groups, with no group showing a dramatic advantage or disadvantage. While there are slight differences, they don't seem substantial without further statistical testing.

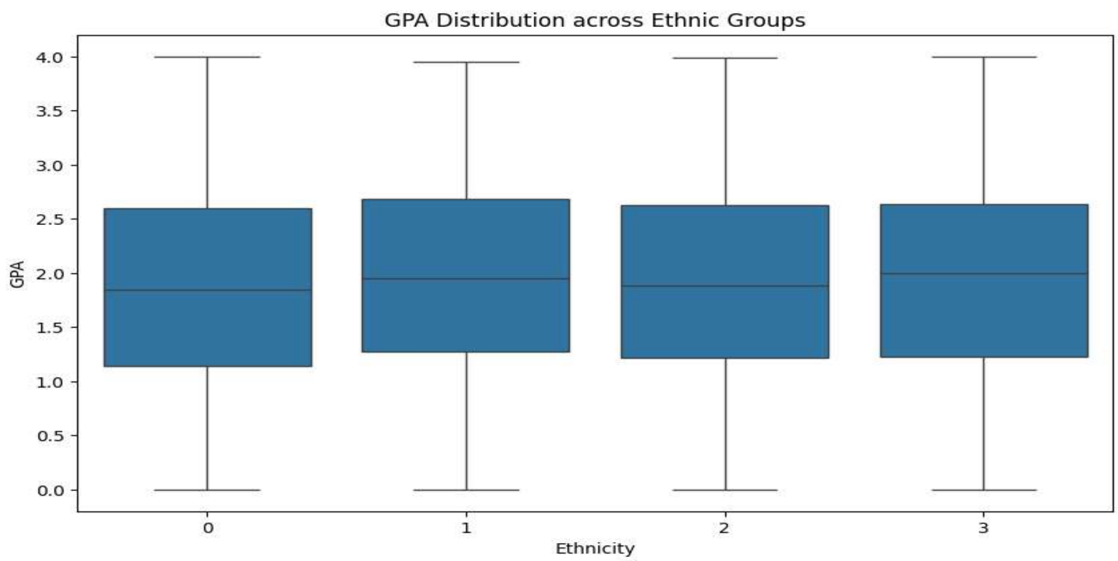
While there are slight differences, they don't seem substantial without further statistical testing
Suspected small differences might not be meaningful — this shows analytical thinking.

Performed a one-way ANOVA test — perfect choice to compare GPA means across multiple groups.
If the p-value < 0.05, there's a statistically significant difference in GPA between ethnic groups.

After Perform one-way ANOVA
F-statistic: 0.96
p-value: 0.4116

Interpretation:

Since the p-value is 0.4116, which is much greater than 0.05, we do not have enough evidence to say that GPA differs significantly across ethnic groups.
F-statistic: 0.96 p-value: 0.4116



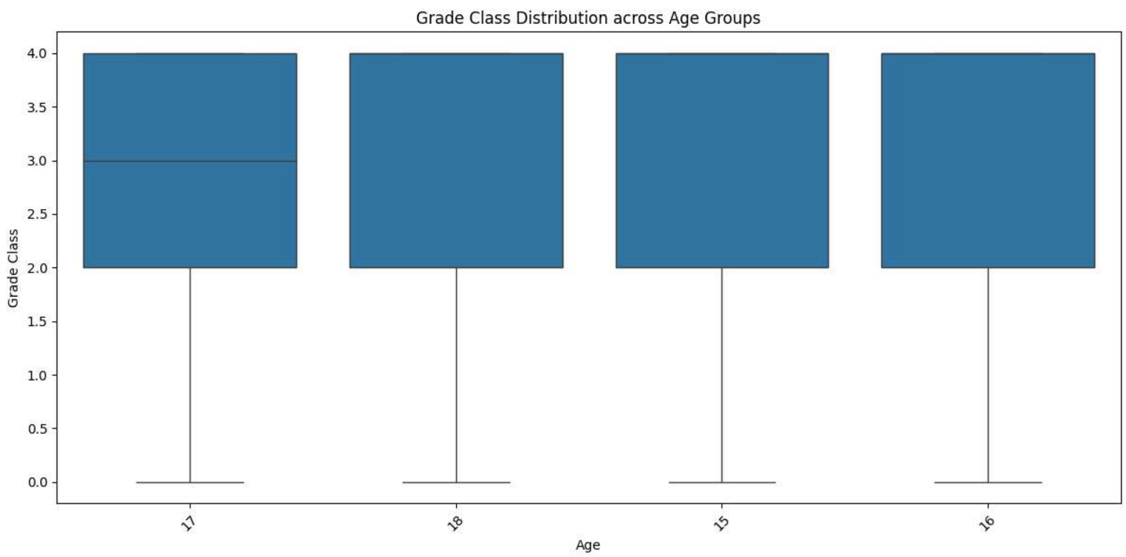
Conclusion:

The small differences in GPA between ethnic groups are not statistically significant.
In simple terms, ethnicity does not appear to have a meaningful effect on GPA in this dataset.

2.10What is the distribution of grades across different age groups?

The boxplot here shows how grade performance, categorized by GradeClass, is distributed across students aged 15 to 18.

- GradeClass 4, which represents failing grades, is **dominant in every age group** — you can see this because the **median line** inside each box is at or very close to 4 for most ages.
- For **ages 15, 16, and 18**, more than **50% of students are in GradeClass 4**, since the **median** is at 4 and the lower half of the box doesn't even extend below that — meaning most students failed.
- •At **age 17**, the median is **lower than 4**, and the box shows more spread into better grades, indicating **a wider and slightly better performance** in this group.



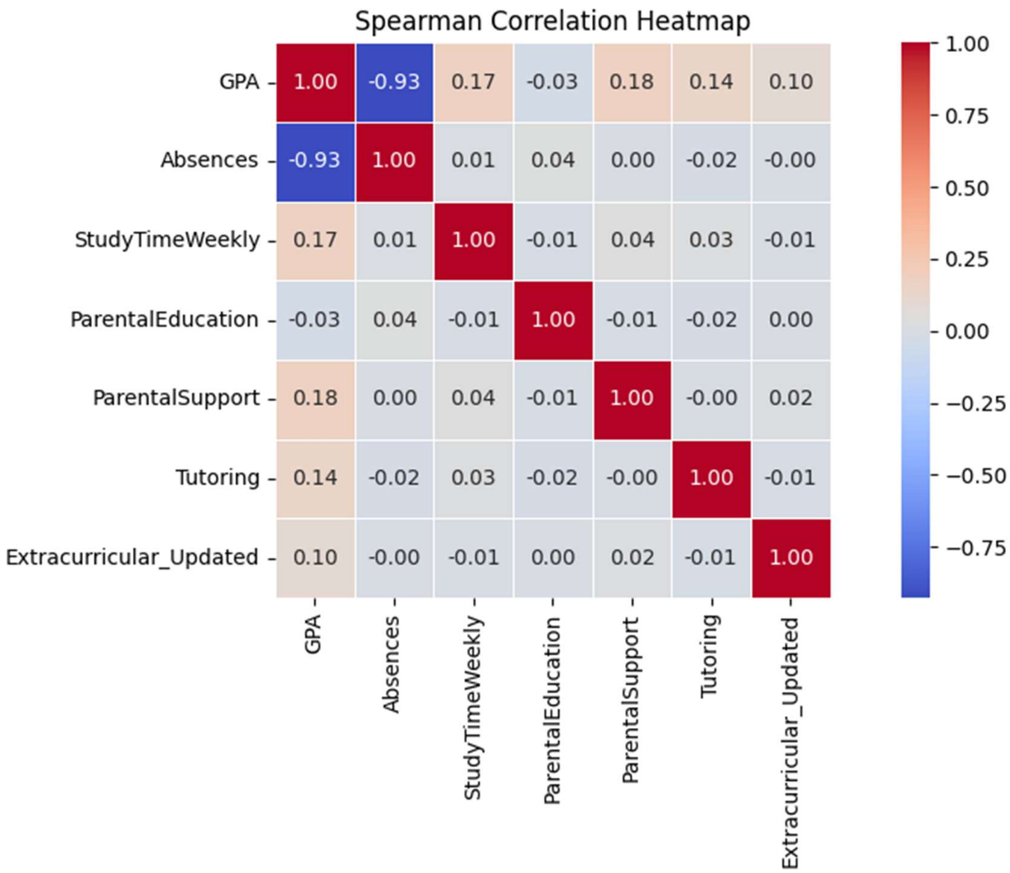
conclusion :

•**Failing grades (GradeClass 4) are most common across all ages**, except age 17.

•**Age 17 students** show **less concentration around failures**, suggesting they perform better than the rest.

Correlation:

This correlation analysis highlights that **attendance (Absences)** is by far the dominant factor related to GPA in a linear sense. Students who attend school regularly perform much better academically. Meanwhile, having supportive parents, dedicating more time to study, receiving tutoring, and participating in extracurricular activities all have positive (though moderate) associations with GPA. Demographic variables (gender, ethnicity, age) and parental education level seem to have little to no direct linear relationship with GPA in this dataset.



Summary of Findings and Conclusions:

Findings: Our analysis reveals that **student attendance (absences)** is the critical factor associated with GPA. Students with high absenteeism almost invariably have low GPAs, whereas students with regular attendance achieve much higher grades. This relationship is so strong that reducing absences should be a top priority for improving academic performance – consistent attendance is fundamental to learning.

Apart from attendance, several other factors significantly influence GPA:

- **Parental Support:** Students who receive strong support from parents (involvement in their education, encouragement, help with schoolwork) tend to perform much better academically. This suggests that engaging parents and guardians in the educational process can yield tangible benefits in student GPA. Schools might consider programs to increase parent-teacher communication, parenting workshops on supporting student learning, and inclusive school events to foster this support. Research backs this up: greater parent involvement is linked to higher achievement and graduation rates.

- **Study Habits (Study Time):** Devoting more hours to studying each week correlates with higher GPA, though with diminishing returns. Encouraging effective study routines – not just quantity but quality of study – is recommended. Since our data indicates a positive but moderate effect, it's important to also focus on *how* students study. Interventions like study skills workshops or monitored study halls could help maximize the benefit of study time.

- **Tutoring:** Access to tutoring (whether through teachers, peers, or external programs) has a clear positive impact on GPA in this dataset. Tutored students on average earned about a quarter-grade higher GPA than non-tutored peers, all else equal. Tutoring can provide personalized attention and address learning gaps, which likely explains the improvement. Schools and districts should consider offering tutoring services or learning centers, especially targeted at at-risk students, to boost their academic outcomes. The investment is justified by substantial returns in performance.

- **Extracurricular Involvement:** Participation in sports, arts, clubs, or other extracurricular activities is associated with higher GPAs. This might be due to improved student engagement, motivation, and time management skills developed through these activities. Students who are involved tend to be more connected to school and might have better discipline balancing their schedule. We recommend not only allowing but encouraging students to pursue at least one extracurricular interest. It's important, however, to monitor that they maintain academic time – but as our analysis shows, involvement generally complements academics rather than competes with it.

- **Demographics (Gender/Ethnicity) and Parental Education:** These factors showed *minimal direct effect* on GPA once the above factors are considered. This suggests that performance differences often attributed to demographics might actually be explained by differences in attendance, support, and resources. In this dataset, boys and girls performed equally, and no ethnic group had an advantage when given similar support and circumstances. Similarly, a parent's education level alone didn't guarantee a student's success – what mattered more was their involvement. This is a positive message: it implies that by improving the modifiable factors (attendance, support, study habits), we can overcome inherent demographic or socio-economic disparities in achievement.

Model Utility: The predictive model developed (a linear regression equation) can be used to estimate a student's GPA with high accuracy. It could serve as an early warning tool: for instance, if a student accumulates many absences or shows low study hours and support, the model would predict a low GPA, flagging the need for intervention. Conversely, the model highlights what changes would most

boost a student's GPA – e.g. reducing absences from, say, 20 to 10 could raise predicted GPA by nearly a full point, far more than the effect of any other single change. This quantitative insight can help prioritize actions.
Recommendations: Based on these findings, the following actions are recommended to improve student GPA outcomes:

1. **Reduce Absenteeism:** Implement strict attendance monitoring and supportive programs to minimize unnecessary absences. This could include attendance incentive programs, parent notifications for frequent absences, dealing with underlying causes (health issues, transportation, school environment) that lead to skipping. Because *"class absenteeism can significantly affect academic performance"*, addressing it will have the most immediate and significant payoff.

2. **Parental Engagement:** Schools should actively involve parents in the education process. Regular updates, PTA meetings focusing on how parents can assist at home, and recognizing parent involvement can maintain high support. For students lacking support, mentoring programs or community "surrogate" support (big brother/sister programs, tutors who also mentor) might help fill the gap.

3. **Academic Support & Tutoring:** Provide accessible tutoring services for students who need extra help. This could be peer tutoring programs, after-school tutoring sessions, or online tutoring resources. Given the evidence that tutoring can improve grades dramatically (often turning failing students into passing ones), scaling up tutoring is a proven strategy to uplift overall GPA and reduce failure rates.

4. **Encourage Effective Study Habits:** Teach study skills and time management as part of the curriculum or through workshops. Encourage students to set aside consistent daily study time. For those who struggle to study at home, offer supervised study sessions at school (e.g., a homework club or library hours). Even modest increases in study time, if done efficiently, can yield improvements.

5. **Promote Extracurricular Participation:** Create or maintain a variety of extracurricular programs (sports, arts, academic clubs) and encourage broad participation. Ensure coaches and club leaders coordinate with teachers so students maintain academic eligibility and balance. Extracurriculars should be seen as an integral part of development – not only do they enrich student experience, they also can improve academic performance by fostering skills that translate into the classroom.

6. **Holistic Interventions:** Recognize that students are benefited by a combination of the above factors. An ideal program might look like: a student with initially low performance is closely mentored (increasing support), attends a homework help/tutoring session daily (increasing study time and tutoring), is tracked for attendance with personalized follow-up for absences (ensuring attendance), and is encouraged to join a club or team to build engagement. This multi-faceted approach targets the key predictors we identified and should result in improved GPA as our analysis predicts.