

**Introduction to Data Science 24CSAI01I**

**Student Performance and Behavior**

Group 7

|  |  |  |
| --- | --- | --- |
| 236295 | Ahmed Hany | [Ahmed236295@bue.edu.eg](mailto:Ahmed236295@bue.edu.eg) |
| 235355 | Amr Galal | [Amr235355@bue.edu.eg](mailto:Amr235355@bue.edu.eg) |
| 232203 | Toka Elsayed | [Toka232203@bue.edu.eg](mailto:Toka232203@bue.edu.eg) |
| 235133 | Habiba Amr | [Habiba235133@bue.edu.eg](mailto:Habiba235133@bue.edu.eg) |
| 232117 | Mennatallah Mohamed | [Mennatallah232117@bue.edu.eg](mailto:Mennatallah232117@bue.edu.eg) |

1. **Description:**

This dataset includes students' personal information, such as student ID, email, name, age, and academic records, such as grades, study hours, attendance, and other performance indicators. It also provides information about parents' education levels and involvement, which is valuable for measuring student performance, predicting grades, and reviewing relationships between studying behaviors and performance. This dataset is helpful for educators, academics, and data analysts who want to improve student performance and academic results.

1. **Questions to be answered by dataset:**

1. How studying hours impact the sleeping hours and the stress levels of the students?

- Compare Sleep Hours and Stress level columns with the grade coloum to see if students with more sleep hours and less stress level get more grades or not.

2. How does internet access affect the project score and student's grades in all departments?

- Check the difference in grades and project scores columns between students with and without home internet access.

3. Does gender and family income play a role in study hours per week and total score?

- Check the gender affection on the number of study hours per week and the total score.

4. How do students’ attendance and study time together influence their participation in extracurricular activities and their total score?

- Compare Attendance percentage and total scores with extracurricular activities (Yes/No) to see if high-attendance students are more involved and get higher scores.

5. How can the family income affect the stress level and the effect of both affect the participation score and total score?

- Check if students with high attendance and more study hours get higher scores.

6. How can the average study hours affect the student's grade in different departments?

- See if departments with more study hours have higher total scores or if some require less studying for high grades.

7. How the total score is affected by either the average quizzes or the projects score upon each department?

- Check if students with high quizzes and assignments scores always get a high total score or not.

8. Do students from higher-income families study more and score higher than those from lower-income families?

- Compare family income level (Low, Medium, High) and check if it affects the study hours per week and total score of students or not.

9. How can family income and parent education level affect choosing a certain department and student's final score?

- Analyze family income level with the Department to see if students from higher-income families choose more expensive or prestigious fields.

10 Does sleeping hours per night affect students' attendance?

- Check if sleep hours per night affect students' attendance or not.

11. What patterns emerge between Grade/mark, Assigned meaning, Final\_Score?

- Analyze the relationship between column Grade/mark, Assigned meaning and Final\_Score.

12. What impact do Family Income Level and GPA have on a student's Assigned Meaning ?

- Check whether Family Income Level and GPA have an impact on a student's Assigned Meaning

13. How do Grade/mark , Assigned meaning, and Internet Access at Home relate to academic performance?

- Check whether Grade/mark , Assigned meaning, and Internet Access at Home relate to academic performance

14. How does student's attendance affect their final grade and GPA in different departments??

- Check whether students with high attendance percentages tend to have higher grades and GPA in different departments or not.

15. How do midterm score, final score and project score affect student's total score?

- Checks the contribution of each one of midterm, final and projects scores in the total score.

16. How do letter grades reflect students’ total scores across different departments?

- Check the effect of the grades over the total score of students across all departments.

17. How does the extracurricular activities and project's score affect the GPA and the precentage of the students?

- Check whether extracurricular activities and project's score affect the GPA and the precentage of the students

**Phase 2:**

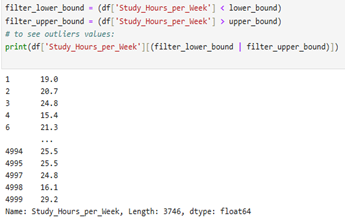
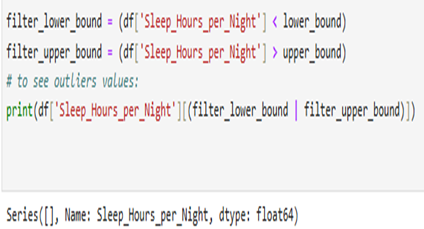
1. How studying hours impact the sleeping hours and the stress levels of the students?

Using scatter plot visualization shows the relationship between study hours, sleep hours, and stress levels. Bubble size for stress level where larger bubbles states higher stress, and Bubble color for sleep hours where fewer sleep hours indicated by cooler colors like blue and more sleep hours indicated by warmer colors like red. The plot shows that students sleeping more are less stressed (small bubbles at top), students studying more and sleeping less have higher stress (large bubbles in lower right). Balancing study and rest can help reduce stress levels.

**Visualization Choice: Scatter Plot**



And by checking outliers in these columns (Sleep Hours Per Night), (Study hours per week), there are no outliers in the (Sleep Hours Per Night) column. But there is for the (Study Hours per week) column.

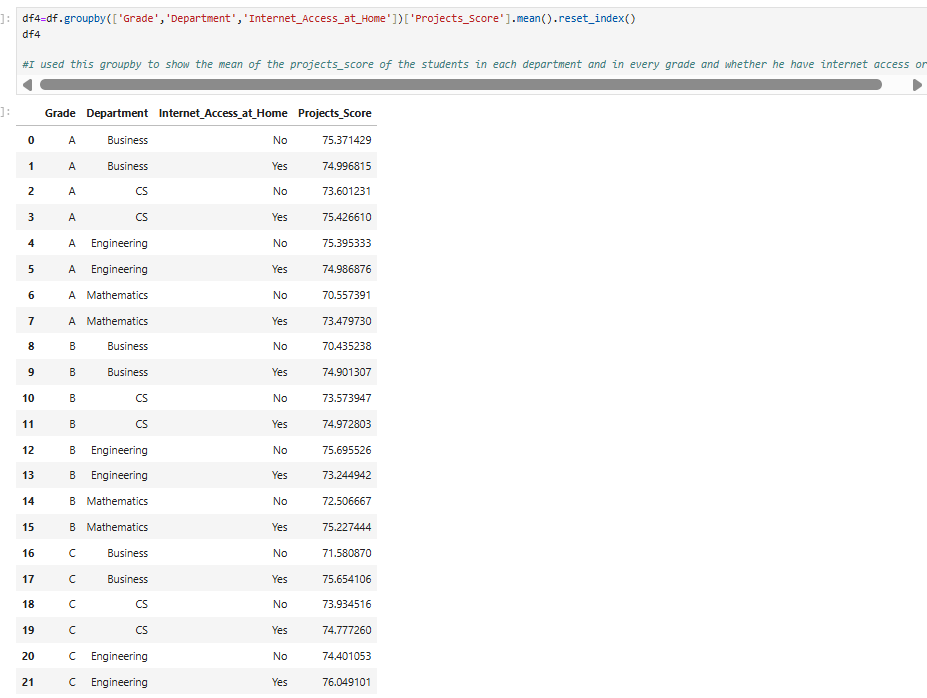
 

2. How does internet access affect the project score and student's grades in all departments?

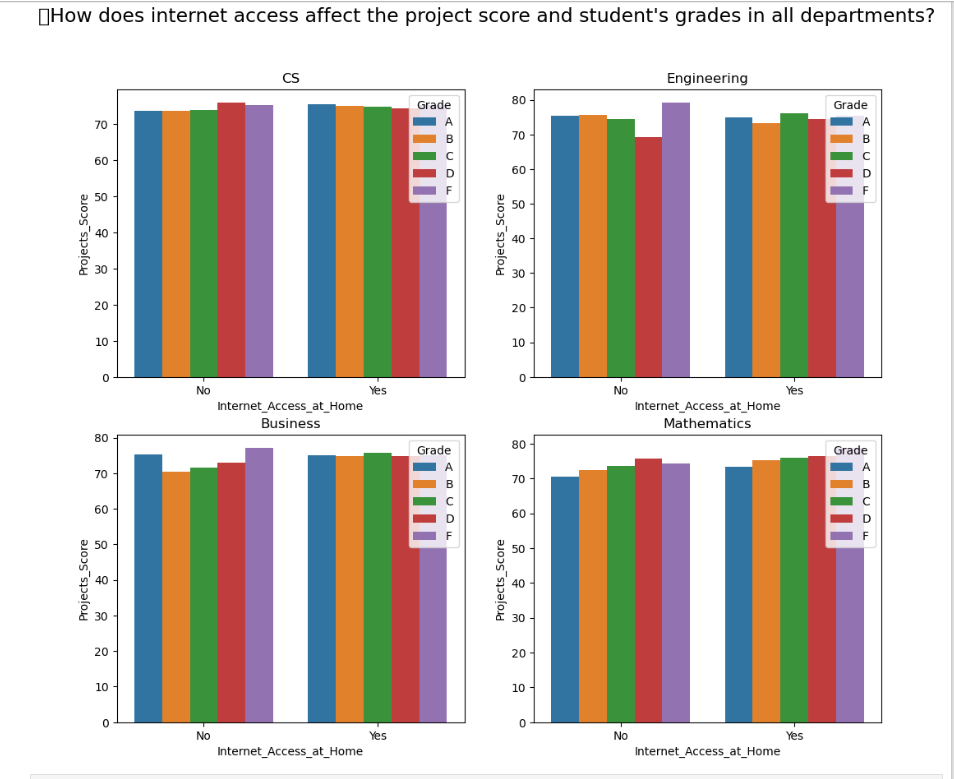
* The question checks if the internet access affect the projects score and student’s grades in each department or not. the best method for visualization is ‘bar plot’, because it clearly compares performance between students with/without internet access in all departments.

**Aggregation:**

* The usage of ‘group by’ here is to show the mean of the projects score of the students in each department and in every grade and whether he have internet access or not.



**Visualization Choice: bar Plot**



**The scatter plot visualizes:**

X-axis: Internet\_Access\_at\_Home

Y-axis: Projects\_Score

hue: Grade

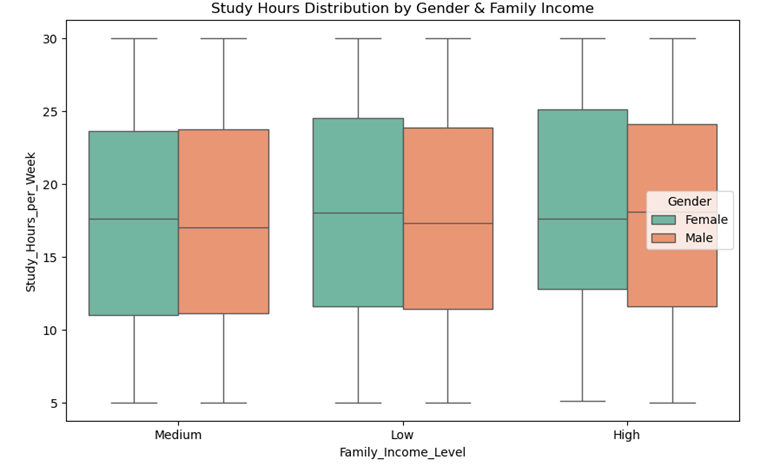
Size: (12,10)

* The plot shows the projects score and student’s grade for the two cases of the internet access at home (yes/no) for the 4 department (CS, Engineering, Business, Mathematics) . The bar plot displays the results and if the internet access affect the projects score and the grades or not.

3. Does gender and family income play a role in study hours per week and total score?

The question checks if the family income level and gender have an impact on study hours and the total score and this determines if the student's gender and their families with income influence their children to have higher score and study more hours per week, which will directly enhance their overall performance. The chosen visualization method is a **box plot**.

**Visualization Choice: box plot**.



**The box plot visualizes:**

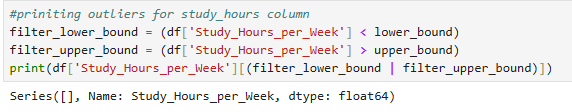
X-axis: Family\_Income\_Level

Y-axis: Study\_Hours\_per\_Week

hue: Gender

Figure size: Adjusts the figure size to 10x6 inches

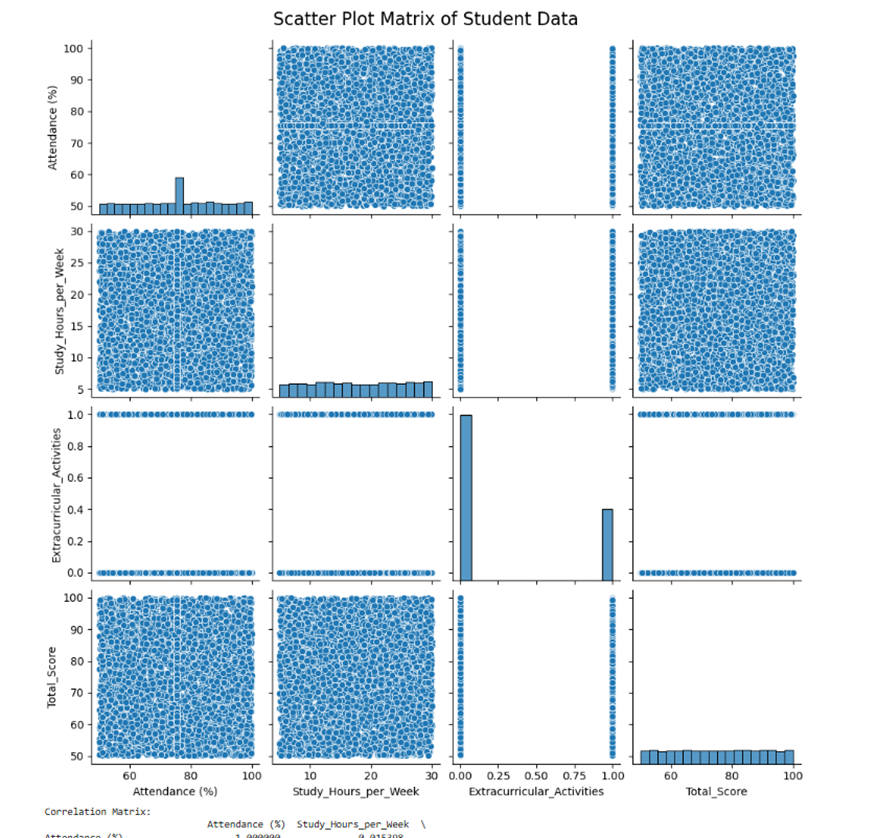
And by checking outliers in the column (Study hours per week), there are no outliers in the column.

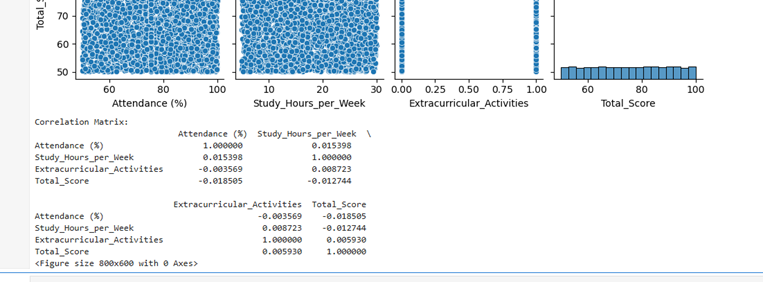


4. How do students’ attendance and study time together influence their participation in extracurricular activities and their total score?

using a pair plot visualization for (Attendance, study hours, Extracurricular activities and Total scores) that has scatter plots between each two variables and histograms on diagonal showing each variable’s distributions. The plot shows that attendance and study hours are positively correlated to total score. Extracurricular activities show small or inconsistent correlation with performance. The pair plot confirms visually what is quantified numerically by the correlation matrix.

**Visualization Choice: pair Plot**





5. How can the family income affect the stress level and the effect of both affect the participation score and total score?

Using aggregation (group by) to compare the average stress level of students among different family income levels (high, low, medium). And showing this relationship (family income affection on stress level) using a bar chart plot(visualization) as the greatest average stress level is for students with medium family income, the lowest average stress level is for students with low family income and the medium average stress level is for students with high family income.

A graph of a family income level

AI-generated content may be incorrect.

Using the four columns (Family Income Level, Stress Level, Participation Score and Total Score) from the dataset to know the affection of family income and stress level on participation and total scores. Showing this using, a pair plot(visualization) indicating that between stress level and participation score data is distributed evenly among stress levels, no clear relation between these two columns that stress doesn’t strongly affect on participation score. Between participation and total scores, the scatter plot between them shows a random pattern that means participation does not determine total score and there is little positive correlation but not strong. And between stress level and total score distribution is random, that means stress does not directly affect total score and total score changes among stress levels. So, this plot shows that stress level doesn’t significantly affect both participation and total scores.

A group of blue bars

AI-generated content may be incorrect.

A group of blue squares

AI-generated content may be incorrect.

And by checking outliers in the three columns (Stress Level, Participation Score and Total Score), there are only outliers in the total score column.

6. How can the average study hours affect the student's grade in different departments?

Using aggregation (group by) to compare grades in each department and the affection of average study hours per week on them, and the ranges of average study hours are slightly different values.

Showing the average study hours impact on grades in different departments using a box plot(visualization), saying that the median of study hours among each grade appears to be similar in among all departments and a wide range for each grade study hours says that study hours doesn’t have a specific grade. Departments have different distributions of study hours that business(orange)students and math(green)students have a slightly taller median study hours, but engineering blue and cs(red) students have comparatively similar distribution. Weirdly, students with lower grades like F and D have some students studying as who have higher grades like A and B. So, there is no strong correlation between grades and study hours as students who study more do not necessarily have better grades.

A graph of different colored bars

AI-generated content may be incorrect.

And by checking outliers in the column (Study hours per week), there are no outliers in the column.

7. How the total score is affected by either the average quizzes or the projects score upon each department?

First Plot:

The question checks if the Projects score affect the total score in each department or not.the best method for visualization is ‘Sub plot using Scatter plot’ , because it explains the relationship between two continuous columns like : Projects\_Score,Total\_Score .

**Visualization Choice: Sub Plot using Scatter Plot**



**The Sub Plot using Scatter Plot visualizes:**

X-axis: Projects\_Score

Y-axis: Total\_Score

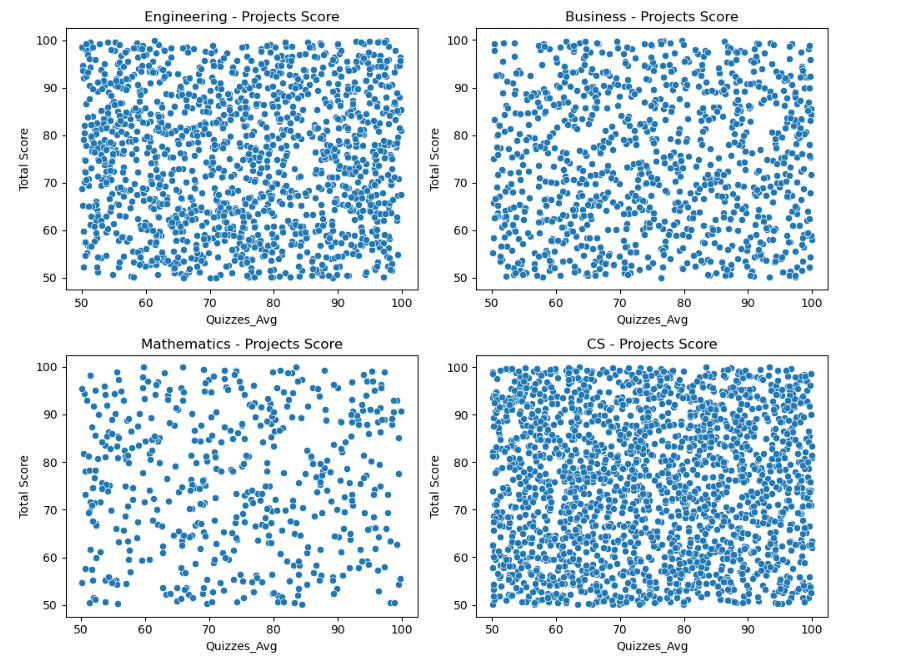
Col: Department

The analysis demonstrates a relationship between project results and complete academic marks in Engineering together with Business and Mathematics and Computer Science. Project performance data fails to show any discernible trend pattern in the four subplots which indicates project scores have no steady connection to academic scorings across the disciplines.

Second Plot:

The question checks if the Quizzes average affect the total score in each department or not. the best method for visualization is ‘Sub plot using Scatter plot’ , because it explains the relationship between two continuous columns like : Quizzes\_Avg,Total\_Score .

**Visualization Choice: Sub Plot using Scatter Plot**



**The Sub Plot using Scatter Plot visualizes:**

X-axis: Quizzes\_Avg

Y-axis: Total\_Score

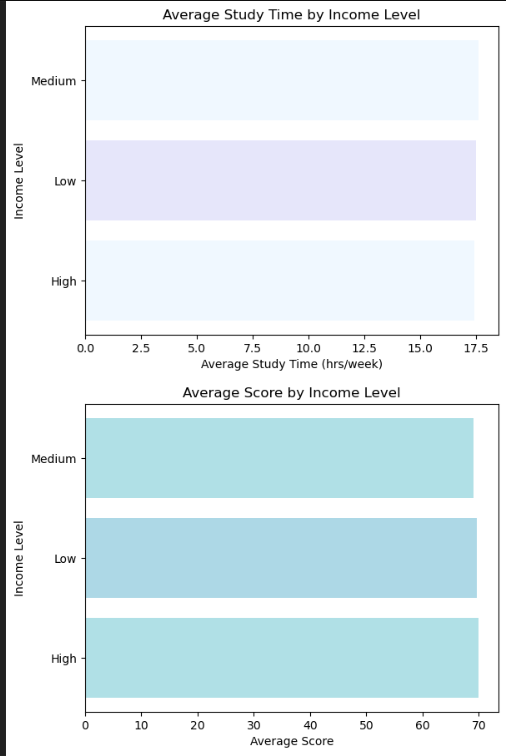
Col: Department

The data points demonstrate how students’ quiz averages relate to their overall scores from four specific majors. Similarly to the project score results the data points show random distribution throughout each plot thus indicating the absence of a strong link between variables. Grade point average is likely unaffected by quiz performance levels based on data analysis.

8. Do students from higher-income families study more and score higher than those from lower-income families?

* The question checks if the family income level has an impact on study hours and the total score and this determines if the families with high impact their children are with the highest score .

**Visualization Choice: Bar Chart**

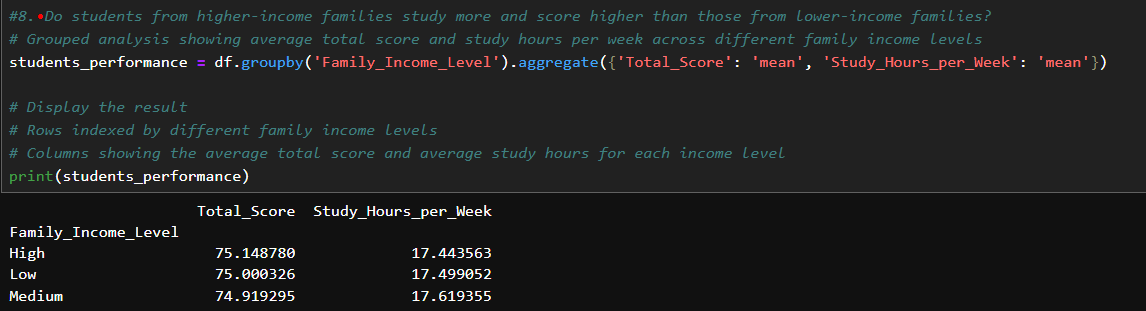
* 

Top Chart:

* Title :Average Study Time by Income Level
* X-axis: Average Study Time in hours per week
* Y-axis: Income Level
* This plot can confirm if higher-income students distribute their study time more extensively than students from lower-income backgrounds. A longer bar for the higher-income students implies they study longer whereas no difference indicates income does not impact study duration.

Bottom Chart:

* **X-axis:** Average Score
* **Y-axis:** Income Level
* Title :Average Score by Income Level
* The plot investigates whether higher family financial resources lead to enhanced grades because bars extending vertically might represent students with better grades and little bar extension might indicate weak performance correlations to family wealth.
* Student performance Aggregation:

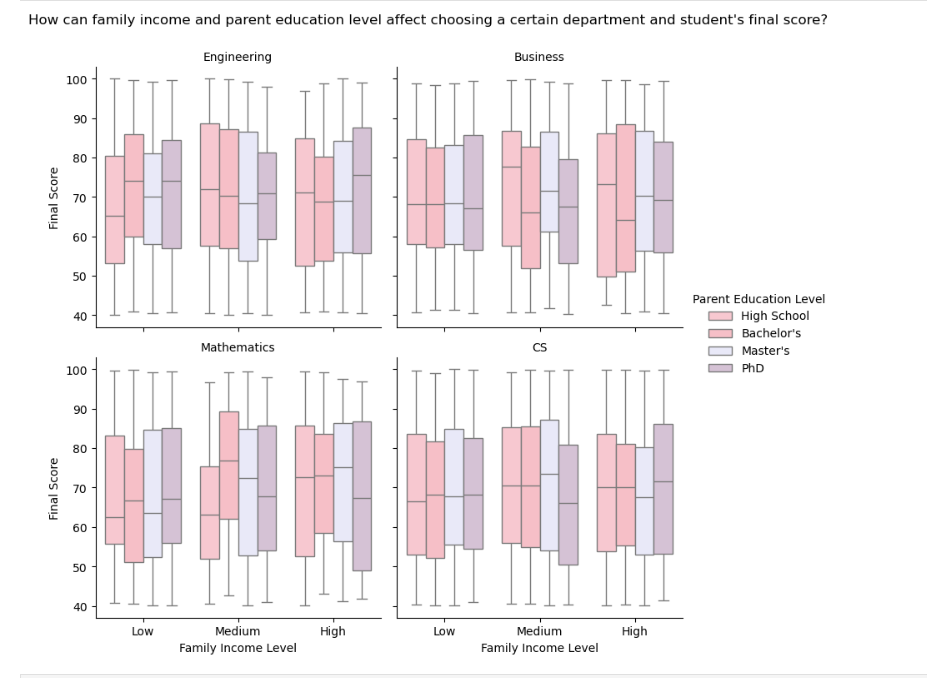


* Groups the Aggregation by the 'Family\_Income\_Level' column .
* Calculates the average (mean) for two columns : Average Total\_Score , Average Total\_Score .
* And then displays the output of the Student performance Aggregation.

9. How can family income and parent education level affect choosing a certain department and student's final score?

This question is analyzing the family income level and parent education level with the Department to see if students from higher-income families choose more expensive or prestigious fields and if the parent education level affect the choices of students for their departments .And if the family income level and parent education level both effect on the total score or not. And the chosen visualization for this question is the scatter plot , because it is one of the most effective plots for clarify values like the family income level , parent education level , department and final score.

**Visualization Choice: Bar Chart**



**The Bar Chart visualizes:**

X-axis: Family\_Income\_Level

Y-axis: Final\_Score

hue: Parent\_Education\_Level

Col: Department

It divides the department into 4 columns with the 4 departments (Engineering , Cs , Business , Mathematics) to see each family income and parent education on each department and if it affect choosing specific department or not and if it affect the final score or not.In our plot, it seems all students perform the same over all department , but some students with parents with higher education tends to have better grades .

10 .Does sleeping hours per night affect students' attendance?

The question checks if there is a relationship between the number of hours students sleep per night and their attendance in school.

**Visualization Choice: box plot**

A screenshot of a computer

AI-generated content may be incorrect.

Title : student performance

X-axis: attendance

Y-axis: sleep hours per night

Rotation : Rotates x-axis labels by 45 degrees to read it clear

Figure size: Adjusts the figure size to 12x6 inches

* Use box plots to compare sleep hours across attendance groups .
* The research examines how sleep patterns affect school attendance because these findings could impact educational results together with student health quality.

The output of the box plot:

A graph with blue and white lines and text

AI-generated content may be incorrect.

The question’s Aggregation:



Groupes the Aggregation by 'Sleep\_Hours\_per\_Night' column.

Calculates the average (mean) attendance percentage for each sleep hours per night group.

**Phase 3:**

1. What impact do Family Income Level and GPA have on a student's Assigned Meaning ?

The question seeks to evaluate how family income and GPA affect students' perceived meaning. A pivot table together with a heatmap visualize the relationship between the investigated variables in this code.



**Visualization Choice: heatmap**

**Creates a pivot table that :**

* Uses GPA as rows
* Uses family\_income\_level as columns
* Counts occurrences of Assigned meaning values in each GPA income combination

Generates a 10x6 heatmap with:

* Color gradient from light yellow (low counts) to dark blue (high counts)
* Title: Number of Students by GPA and Family Income Level

X-axis: Family Income Level

Y-axis: GPA

Academic GPA and household income levels show their connection patterns by the visual data presentation in the heatmap. The intensity of blue color within cells shows the areas where various income ranges combine with different GPA ranges. The chart allows readers to comprehend family income effects on student academic performance and identify which regions need the highest amount of assistance.

2. How do Grade/mark , Assigned meaning, and Internet Access at Home relate to academic performance?

A comprehensive examination explores three major influencing factors between grades/marks, assigned meaning and home internet access which affect academic performance. A structured breakdown follows of the investigated relationships with their potential academic performance effects.



**Visualization Choice: Bar Chart**

Groups data by three key variables:

Grade/mark ,assigned meaning, internet access at home

Title: Academic Performance by Grade, Assigned Meaning & Internet Access

X-axis: grade /mark

Y-axis: average total score

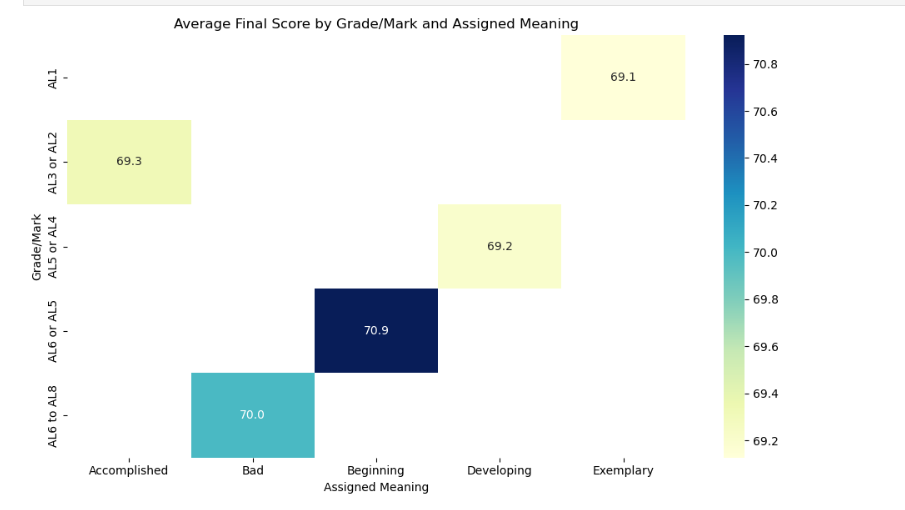
Hue: internet access

The bar plot shows that students who have internet access at home continuously perform better in school especially at advanced academic levels. Students who have digital connectivity tend to experience better academic achievements particularly when they demonstrate advanced levels of understanding. Foundational academic skills seem to require less dependence on internet compared to higher levels. The investigation requires further research to examine the quality factors of internet usage.

3. What patterns emerge between Grade/mark, Assigned meaning, Final\_Score?

This question is analyzing the Grade/Mark and Assigned meaning with the Final score to see what is the relationship between those columns ,how it affect each other and affect the student performance .The chosen visualization for this question is 'Heat Map', because it is one of the most effective plots for clarify values like the Grade/mark , Assigned meaning , Final score.

**Visualization Choice: Heat Map**



**The Heat map visualizes:**

X-axis: Assigned meaning

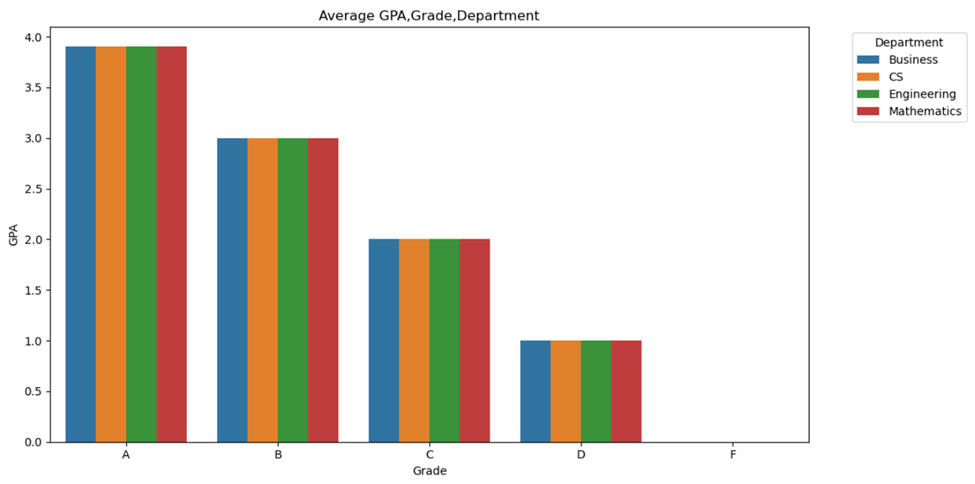
Y-axis: Grade/Mark

hue: Fina\_Score

Student average final scores appear in this heatmap through a two-dimensional display of their grade level on the vertical axis and their assigned meaning level on the horizontal axis. The colored cells demonstrate different combinations of grade level and assigned meaning which show average final scores through numbers inside them. The highest recorded value of 70.9 exists in the group classified as "Beginning" to "A16 or A15." Contrary to most other measures which clustered near 70. Score difference is shown through color intensity however categories contain minimal variations since assignment meanings and grade levels fail to significantly affect average performance.

4. How does student's attendance affect their final grade and GPA in different departments??

Using aggregation group by to show the mean/average GPA for each Grade in different Departments. Using a bar plot visualization where every bar group is for different grades (A, B, C, D or F), and every colored bar in each group is for different departments (Business, CS, Engineering and Mathematics). And the plot shows that higher grades students have higher GPAs, GPA uniformly decreases with low grades, and there is no big difference between different departments, then GPA trends by the grade are the same in different departments.



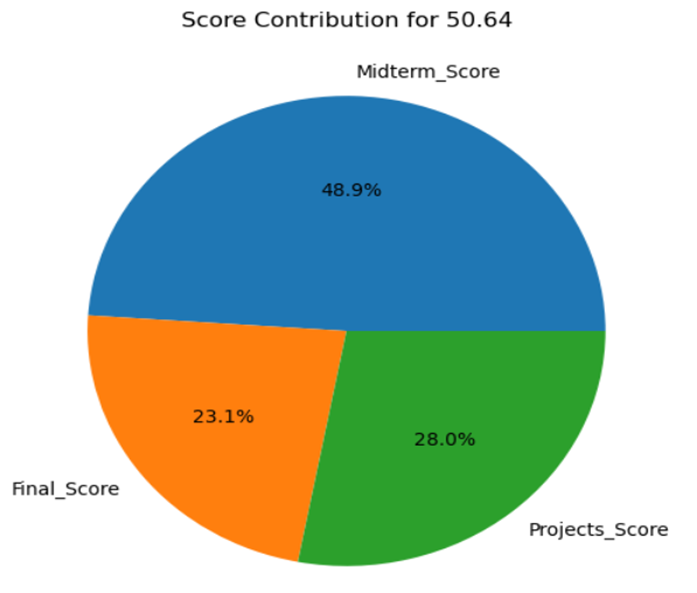
Using a scatter plot visualization where each point is a student and points are color-coded by the grades, attendance percentage on x-axis and GPA on y-axis. And the plot shows that students with higher attendance have higher GPAs, students with high grades like (A and B) are placed at the upper right side which means high attendance and GPA, and students with low grades like (D and F) are placed at the lower left side which means low attendance and GPA.



And by checking outliers in the columns (GPA and Attendance), it has outliers only in the GPA column checked in the scatter plot.

5. How do midterm score, final score and project score affect student's total score?

Using pie plot visualization shows the contribution of each one of the midterms, final and project scores of the first student in the data set to the total score. The total score is 50.64 and contribution of midterm score is 48.9%, projects score 28.0% and final score 23.1%. which means that midterm scores strongly influence on total score, project contribution is moderate, and final score had the lowest impact. Then, the student should prioritize performing well on midterm, followed by projects, then final in order to maximize total score.

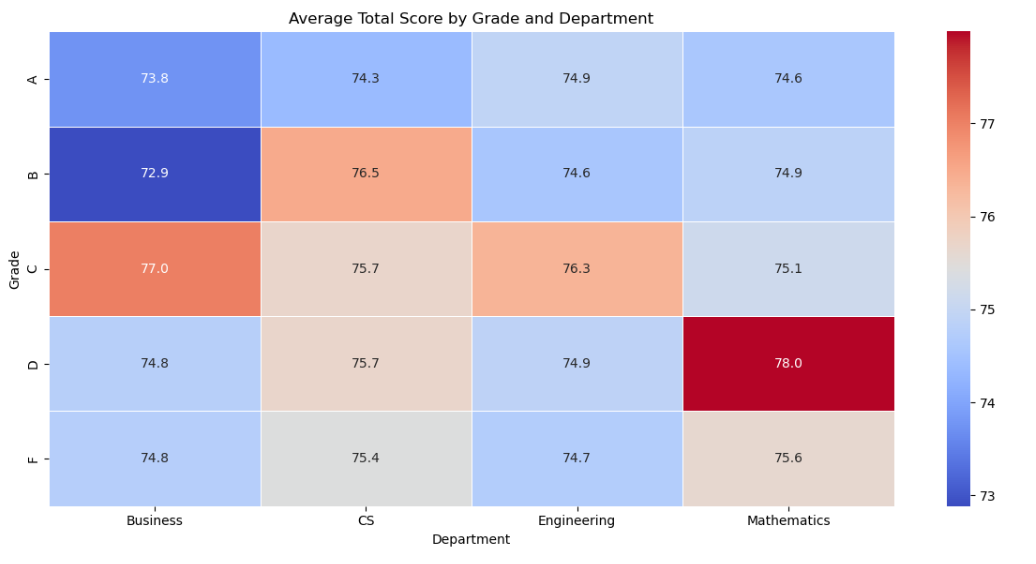


And by checking outliers in the columns (Midterm Score, Projects Score, Final Score and Total Score), there are no outliers in them.

6. How do letter grades reflect students’ total scores across different departments?

This question is analyzing the Grades of the students and compare it to the total score of students in each department to see if it has affected or not .And the chosen visualization for this question is the ‘Heat map’ , because it is one of the most effective plots for clarify the effective of these column in a very clear way.

**Visualization Choice: Heat Map**



**The Heat map visualizes:**

X-axis: Department

Y-axis: Grade

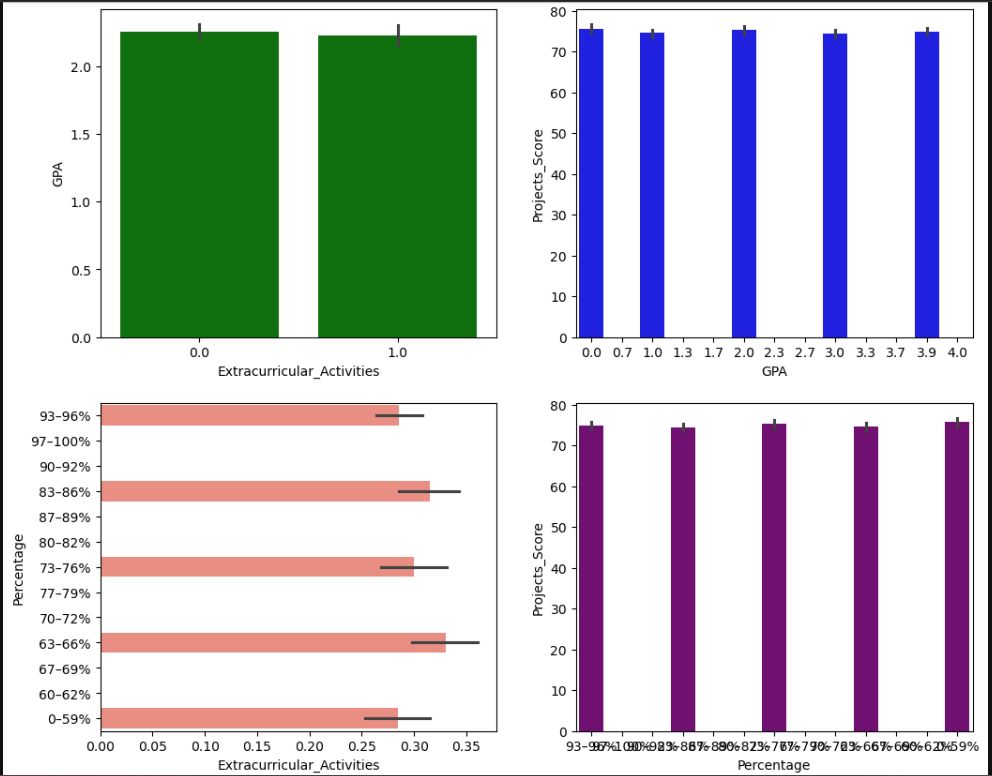
hue: Total\_Score

Based on heatmap analysis, student achievement stands as the main factor affecting final department score levels . The overall scores of students increase through a transition from lower grade levels to upper grade levels in Engineering and Science departments. The department scores at Arts and Business show unstable performance patterns. Student performance at school depends on both their major department and their educational year.

7. How does the extracurricular activities and project's score affect the GPA and the precentage of the students?

Four subplots within the visualization study how students' grades and percentage scores associate with GPA along with project scores and involvement in extracurricular activities. Each subplot within the visualization provides important details which shape the report's interpretation as explained below.

**Visualization Choice: Bar Plot**



1. Extracurricular Activities and GPA

A bar plot comparing the average GPA who participate in extracurricular activities.

2.GPA and projects score

A bar plot showing the relationship between GPA levels and project score.

3. extracurricular activities and percentage

A bar plot comparing the average percentage scores of students based on extracurricular activities.

4.percentage and projects score

A bar plot examining how percentage scores relate to project scores .

* Academics improves noticeably when students participate in extracurricular activities because they boost their GPA scores as well as their percentage scores and project outcomes. Educational institutions working with teachers need to implement activities which advance students toward comprehensive growth.
* Academic data from GPA and percentage reveal a strong relationship that shows strong performers in one area will also demonstrate superior performance in projects.

**Conclusion:**

In conclusion, we used aggregations and different visualizations (plots) to answer 10 questions of phase 1 about our student’s performance and behavior dataset, and 7 questions of phase 2. We used different visualizations including (scatter plot, bar plot, box plot, pair plot, subplot, heatmap and pie plot) to show the relations, correlations and trends between two or more columns from the dataset. Aggregations used to get mean/averages of numerical columns. From the important relations answered in the questions are the studying hours impact on sleeping hours and stress levels, attendance and study hours impact on total score and extracurricular activities, family income and stress levels affecting participation and total scores, parent’s education level and income effect on choosing department and final score, assignment averages, GPA and assigned meaning relationships in departments, Grade/mark and assigned meaning relation to performance, attendance affecting grade and GPA, and finally, total score affected by midterm, final and projects scores. Overall, our analysis shows relationships and hidden patterns and trends in our dataset by showing plots relating to different columns.