

Assignment1(Individual/ Group of two)
CS160
Introduction to Data Science
Fall 2023

Working on Techniques for Analyzing Data

Instructions: Complete the following activities for this project.

1. Create a new GitHub repository named Assignment1_XXX, where XXX are your initials.
2. Using excel (to generate the result) and word documents (type answers and paste the results) work on the following questions and submit your work using **pdf** format.

Description:

This dataset contains information about exam scores of a group of students. It includes attributes such as student ID, gender, age, subject, exam score, and study hours.

Attributes:

Student ID: A unique identifier for each student.

Gender: The gender of the student (male or female).

Age: The age of the student.

Subject: The subject of the exam (e.g., Math, Science, English).

Exam Score: The score achieved by the student in the exam.

Study Hours: The number of hours the student studied for the exam.

Objective:

Perform a descriptive analysis of the student exam scores to understand factors affecting performance and identify trends.

- A. **Summary Statistics:** Calculate summary statistics for exam scores and study hours (mean, median, standard deviation, etc.).

<i>Exam Scores</i>		<i>Study Hours</i>	
Mean	85.01111	Mean	4.466667
Standard Error	0.726955	Standard Error	0.120548
Median	86	Median	4
Mode	88	Mode	4
Standard Deviation	6.896497	Standard Deviation	1.143619
Sample Variance	47.56167	Sample Variance	1.307865
Kurtosis	-0.76854	Kurtosis	-1.25364
Skewness	-0.3694	Skewness	-0.03155

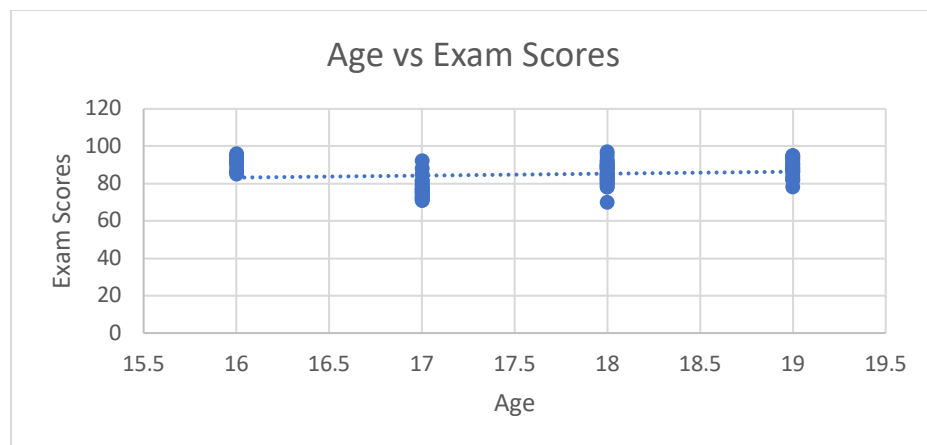
Range	27	Range	4
Minimum	70	Minimum	2
Maximum	97	Maximum	6
Sum	7651	Sum	402
Count	90	Count	90

B. Gender Analysis: Compare average exam scores and study hours for male and female students using PivotTables or simple calculations.

Row Labels	Average of Exam Score	Average of Study Hours
English	83	4
Female	87	5
Male	80	4
Math	86	5
Female	90	5
Male	82	4
Science	86	5
Female	91	5
Male	80	4
Grand Total	85	4

The average exam scores for females were much higher in all three subjects than the average exam scores for males. Both genders studied for several hours on average, but females had a higher average of study hours overall.

C. Age Analysis: Analyze how exam scores vary with age using scatter plots or trend lines.



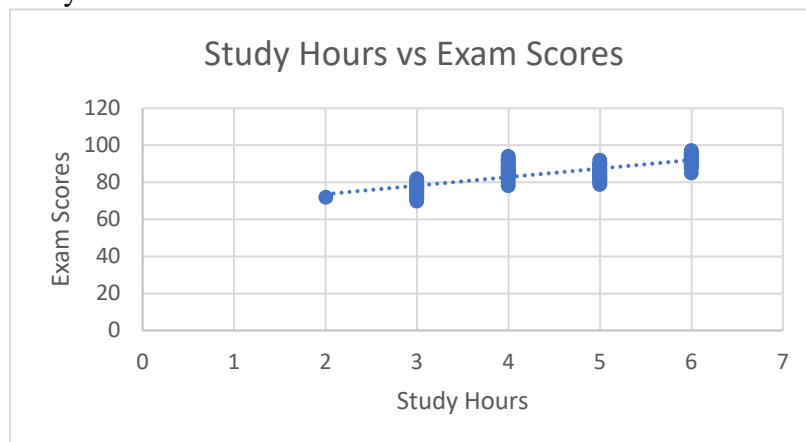
The exam scores and age have a weak relationship. Age does not have a large impact on exam scores and is proven by the correlation of .15, which signifies a weak correlation. Exam scores are not dependent on age.

- D. **Subject Analysis:** Explore average scores for each subject to identify strengths and weaknesses.

Row Labels	Average of Exam Score
English	83
Math	86
Science	86
Grand Total	85

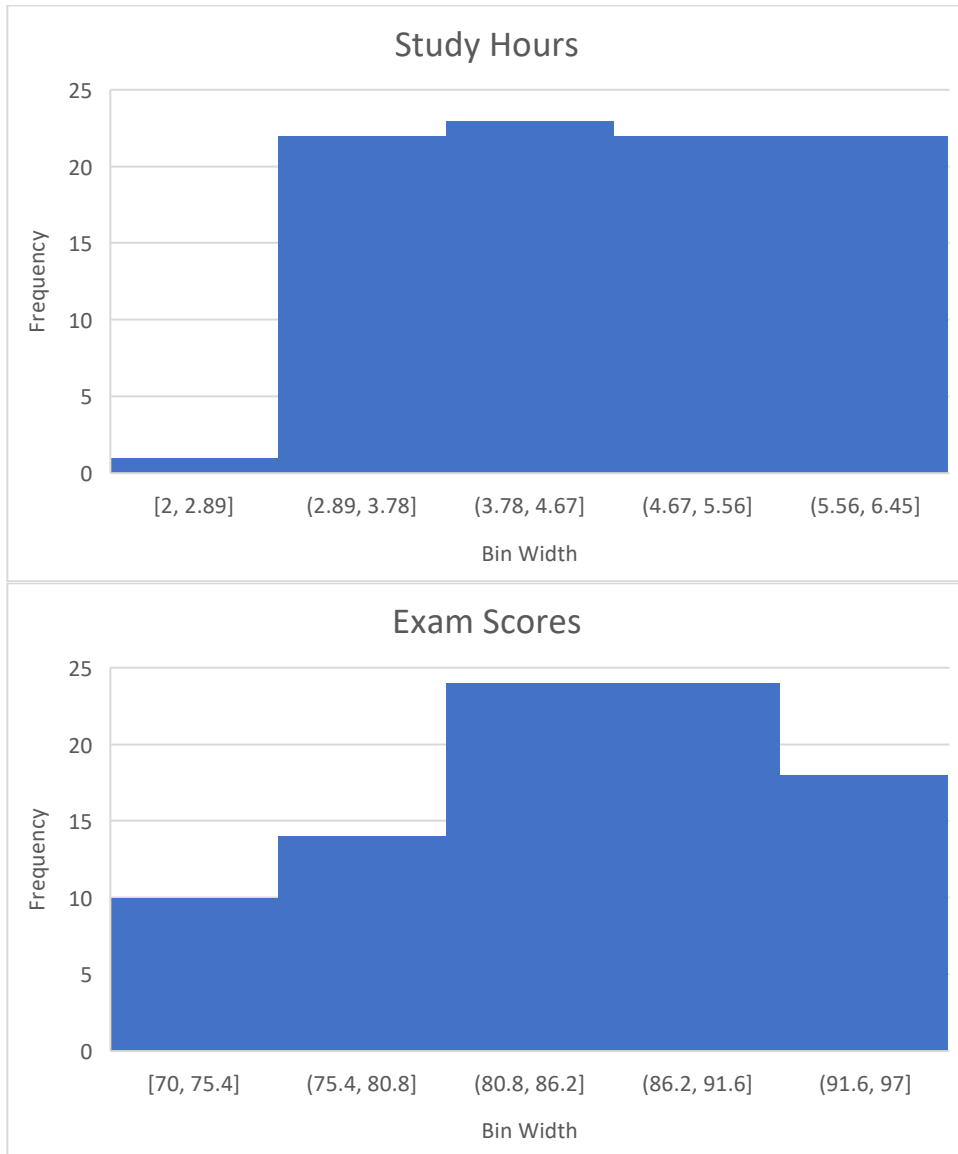
Although the average scores are very similar, English would be the weakness because it is the lowest average exam score. Math and Science are the strongest subjects because the students had the highest average exam scores in these classes.

- E. **Study Hours vs. Exam Score:** Create a scatter plot to visualize the relationship between study hours and exam scores.



There is a positive correlation between study hours and exam scores. We can determine that the amount of study hours that a student has will have an impact on their exam score. The trend line shows that there is a positive relationship, and the exact correlation value is .764, which shows that there is a strong, positive relationship between study hours and exam scores.

- F. **Distribution Analysis:** Create histograms to show the distribution of exam scores and study hours.



The distribution of exam scores is slightly skewed to the left and is not completely symmetrical. Study hours is a fairly uniform distribution with an outlier in the 2-2.89 range.

G. **Top Performers:** Identify students with the highest scores and analyze their study hours, gender, and age.

The top 10 students are all females. Additionally, they are all extremely successful in either Math or Science, with only one of the students in the top 10 having their highest score in English. Two of the students are 16 years old, with the rest of them being in the

age range of 18-19. The correlation between gender and exam score cannot be determined by the CPU, but it is significant that the top 10 students are all females.

- H. **Correlation Analysis:** Calculate the correlation between study hours and exam scores to understand their relationship.

The correlation between study hours and exam scores is .764358. This indicates that there is a strong, positive relationship between study hours and exam scores. Exam scores are dependent on the amount of study hours that a student endures. The more a student studies, the better they will score on the exam.

3. Provide a summary result for each of the findings.

In conclusion, exam scores are most dependent on study hours. Age does not have a large impact on exam scores. The strongest correlation is between study hours and exam scores, which means that exam scores are most dependent on study hours.

4. Using the instructions provided by GitHub, create a git repository named **DS160InClassAssignment**, and push your pdf file to it. Each of you needs to submit your work.

Submission:

Paste a link to your GitHub repository in the area provided for this assignment and submit it by class time.