

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	85.01111111	Mean	4.466667
Standard Error	0.726954629	Standard Error	0.120548
Median	86	Median	4
Mode	88	Mode	4
Standard Deviation	6.896497148	Standard Deviation	1.143619
Sample Variance	47.56167291	Sample Variance	1.307865
Kurtosis	-0.768538292	Kurtosis	-1.25364
Skewness	-0.369397246	Skewness	-0.03155

Range	27	Range	4
Minimum	70	Minimum	2
Maximum	97	Maximum	6
Sum	7651	Sum	402
Count	90	Count	90
<hr/>			
	1		1

Exam Score:

Mean- 85.01111111

Median- 86

SD- 6.896497148 Study

Hours:

Mean- 4.466667

Median- 4

Mode- 4

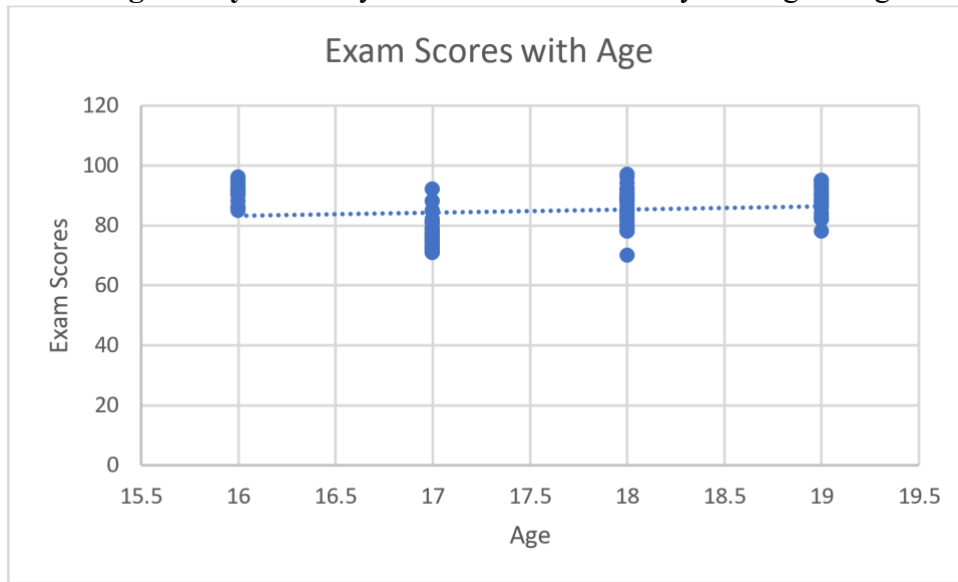
The average of hours studied was 4 while the median of test scores was 86.

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
<u>English</u>	<u>83</u>
Female	87
Male	80
<u>Math</u>	<u>86</u>
Female	90
Male	82
<u>Science</u>	<u>86</u>
Female	91
Male	80
Grand Total	85

Females did better than males in all subjects (English, Math, and Science).

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



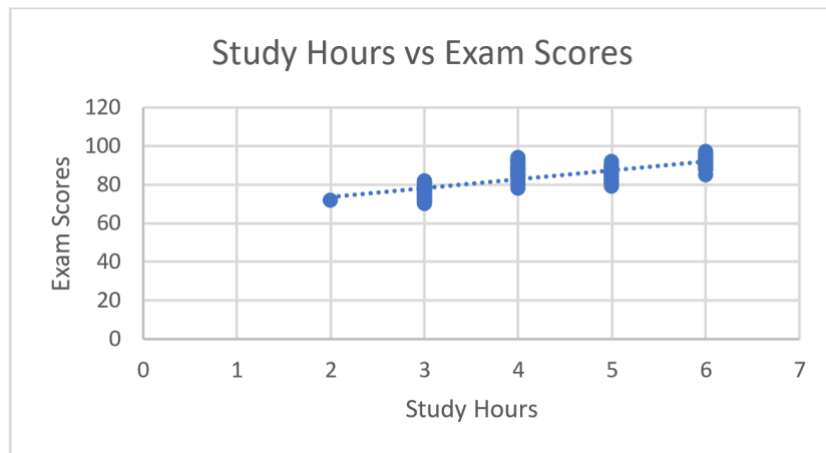
From this trend line, the correlation is 0.150218. It shows that there is a weak relationship of exam scores and age.

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

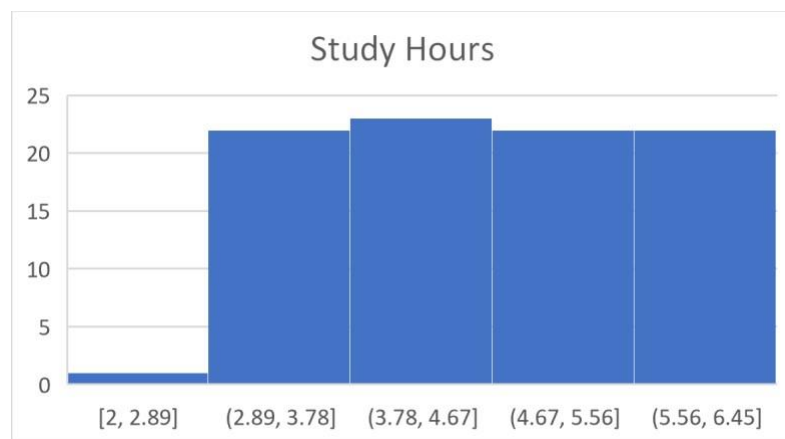
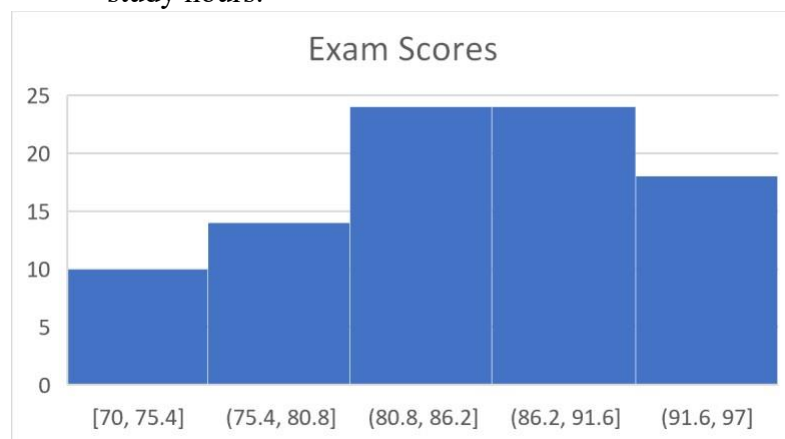
Students performed a little better in STEM courses than the English subject.

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



Correlation: 0.764358. It shows that the more hours someone studies the higher the exam scores.

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



From the two histograms, you can see there is a right-skewed of the exam scores while the number of study hours were similar however there seems to be an outlier between 2 to 2.89 hours.

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

While analyzing the top 10 performers, more students studied from 5.2 to 6.4 hours. All of the top 10 performers were all female and the age ranged from 16-19 years old.

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

Correlation: 0.764358

There is a pretty strong relationship between the number of study of hours determining exam scores.

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

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.