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

<div align="center">

Working on Techniques for Analyzing Data

</div>

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.).

Exam Score		Study Hours	
Mean	85.01	Mean	4.47
Standard Error	0.73	Standard Error	0.12
Median	86.00	Median	4.00
Mode	88.00	Mode	4.00
Standard Deviation	6.90	Standard Deviation	1.14
Sample Variance	47.56	Sample Variance	1.31
Kurtosis	-0.77	Kurtosis	-1.25
Skewness	-0.37	Skewness	-0.03

| Range	27.00	Range	4.00
Minimum	70.00	Minimum	2.00
Maximum	97.00	Maximum	6.00
Sum	7651.00	Sum	402.00
Count	90.00	Count	90.00

On average, students studied for almost 4 ½ hours and scored 86% on their exam scores.

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

Row Labels	Average of Study Hours	Average of Exam Score
Female	4.96	89.36
Male	3.98	80.67
Grand Total	4.47	85.01

Females tended to study an hour more than males and averaged higher exam scores by 9%.

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

Exam Scores by Age
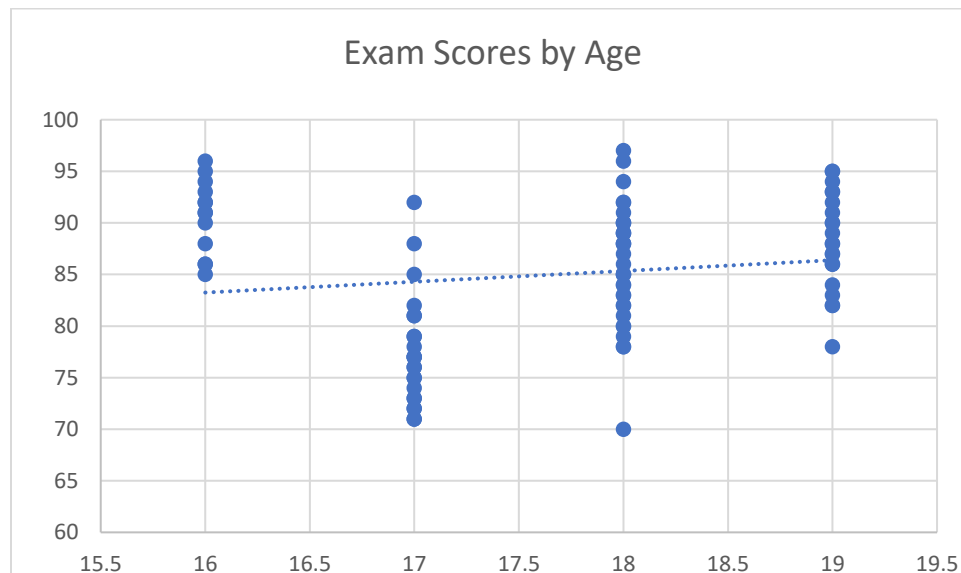
There is no correlation between age and exam scores, with the correlation being at a low 0.15. The trendline is closer to having no slope, relating to the lack of correlation. Exam score is not dependent on age.

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

D. **Study Hours vs. Exam Score:** Create a scatter plot to visualize the relationship between study hours and exam scores.
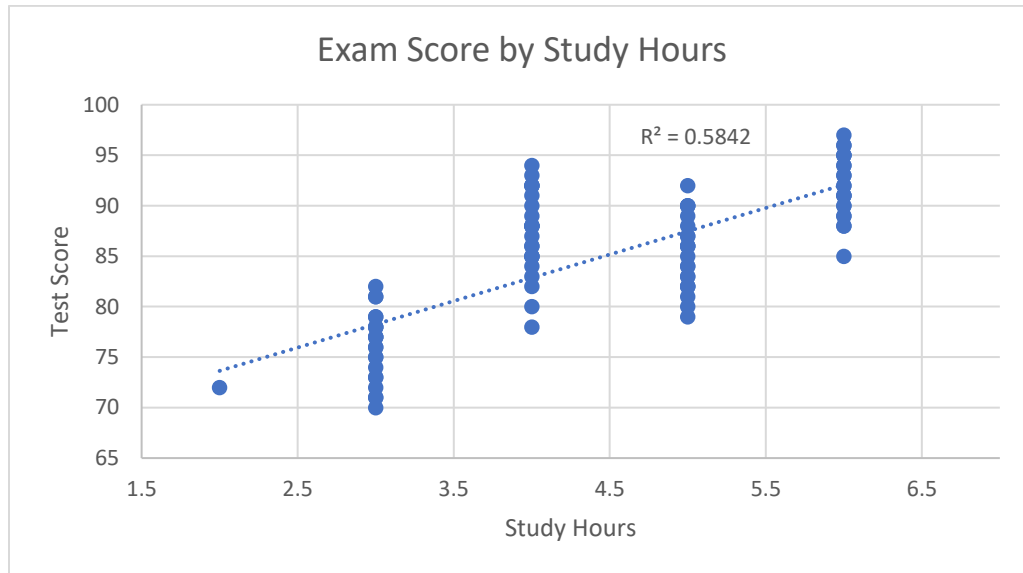
Exam scores are dependent on study hours. With a positive linear relationship and a correlation of 0.76, it reflects that students who study longer are more likely to score higher on their exams.

E. **Distribution Analysis:** Create histograms to show the distribution of exam scores and study hours.
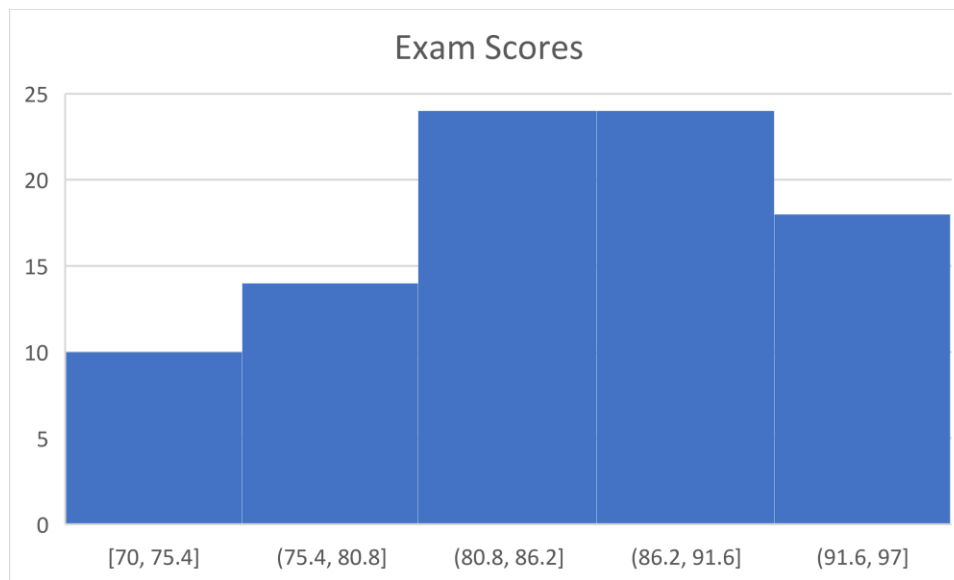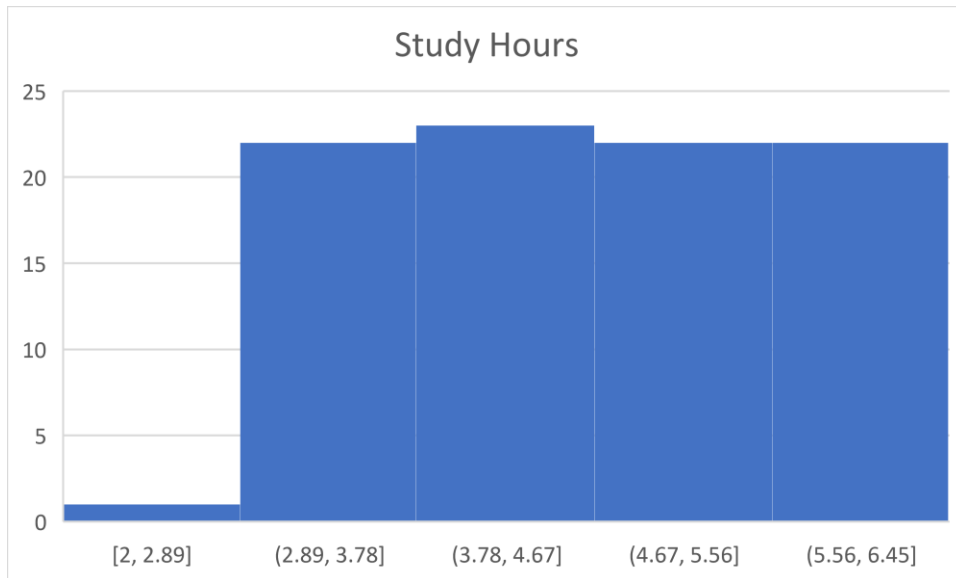
Study Hours

Exam scores are slightly skewed to the left. Study hours is more uniform than exam scores, but has includes an outlier.

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

Gender	Age	Subject	Exam Score	Study Hours
Female	18	Science	97	6
Female	16	Science	96	6
Female	18	Science	96	6
Female	16	Math	95	6
Female	19	Math	95	6
Female	19	Math	95	6
Female	18	Science	94	6
Female	16	Math	94	4
Female	19	Math	94	6
Female	19	Math	93	6
Female	16	English	93	4
Female	19	Science	93	6

All top 10 exam scores are female, science holds most of the highest scores. The max exam score was a 97 and minimum in the top was 93. Most studied 6 hours.

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

The correlation between study hours and exam scores is 0.76. While the relationship isn't the strongest, a positive linear one is there… meaning students who study more are more likely to score higher on their exams.

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

4. Using the instructions provided by GitHub, create a git repository named DS160**InClassAssignment**, 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.