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.).

Exam Score		*Study Hours*	
Mean	85.0	Mean	4.5
Standard Error	0.7	Standard Error	0.1
Median	86.0	Median	4.0
Mode	88.0	Mode	4.0
Standard Deviation	6.9	Standard Deviation	1.1
Sample Variance	47.6	Sample Variance	1.3
Kurtosis	-0.8	Kurtosis	-1.3
Skewness	-0.4	Skewness	0.0
Range	27.0	Range	4.0

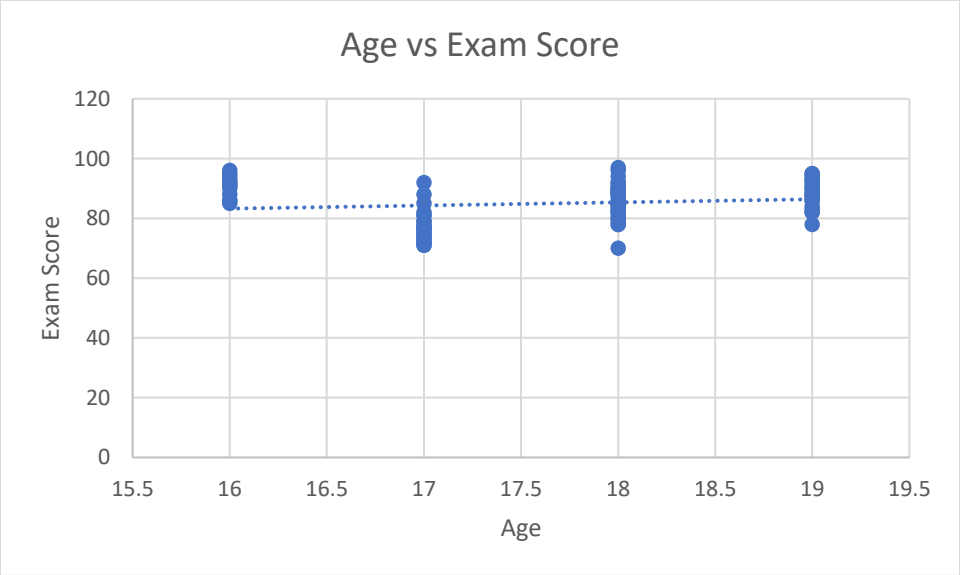
Minimum	70.0	Minimum	2.0
Maximum	97.0	Maximum	6.0
Sum	7651.0	Sum	402.0
Count	90.0	Count	90.0
Confidence Level(95.0%)	1.4	Confidence Level(95.0%)	0.2

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	Sum of Study Hours
English	**83**	**124**
Female	87	66
Male	80	58
Math	**86**	**142**
Female	90	76
Male	82	66
Science	**86**	**136**
Female	91	81
Male	80	55
Grand Total	**85**	**402**

The female students tended to study more and achieve higher scores.

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



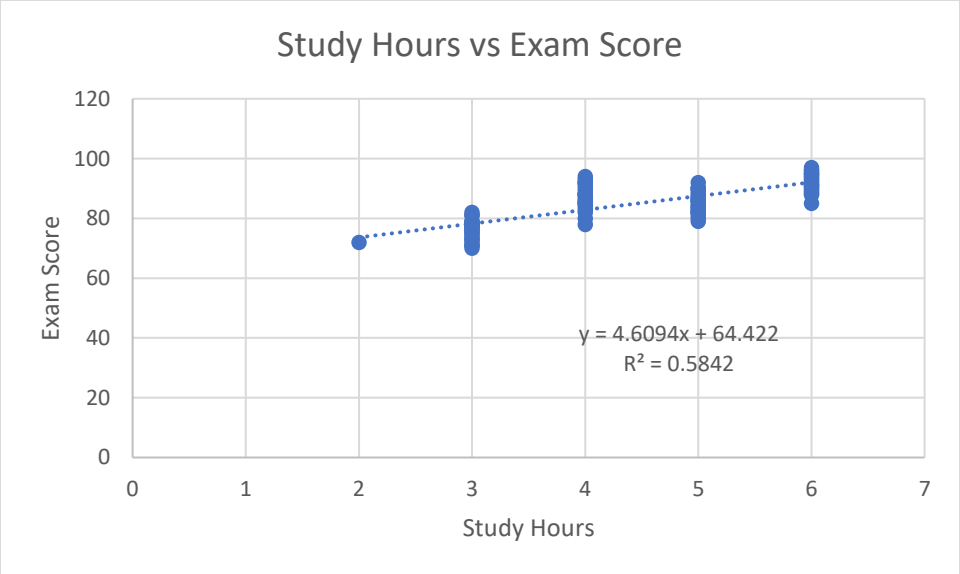
Exam score is **not** dependent on age.

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**

Math and science are the stronger subjects, while more students struggled with the English.

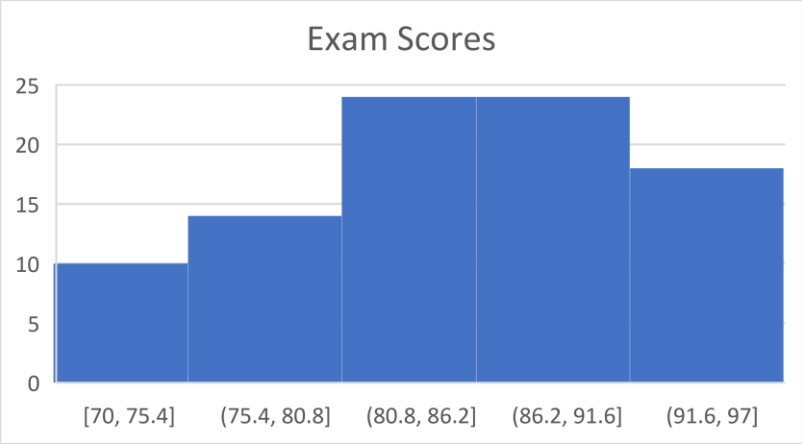
E. **Study Hours vs. Exam Score:** Create a scatter plot to visualize the relationship between study hours and exam scores.



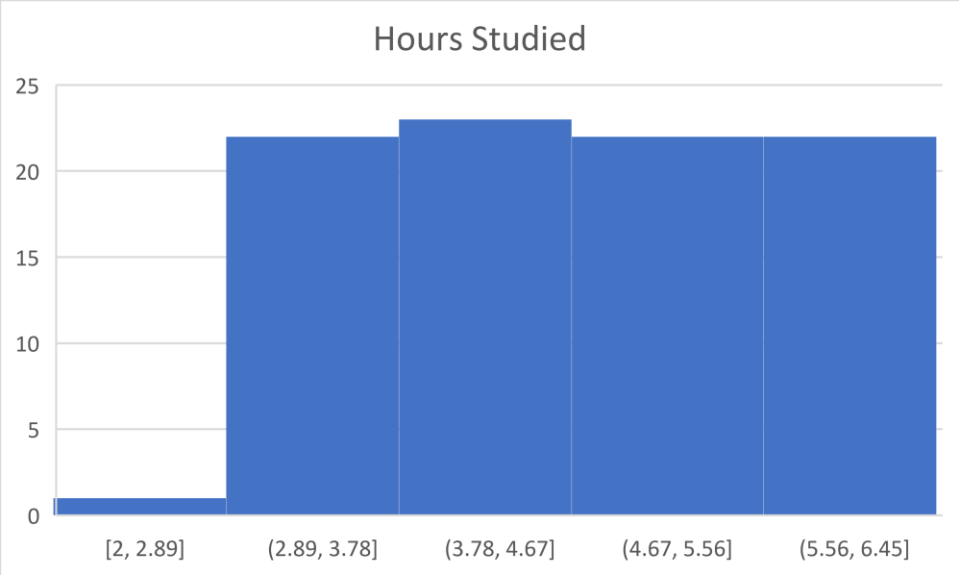
Correlation = 0.764358

There is a positive correlation between more study hours and higher exam scores

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



The exam scores are slightly skewed to the left



The hours studied are more uniform than the exam scores.

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

Student ID	Gender	Age	Exam Score	Study Hours
90	Female	18	97	6
8	Female	16	96	6
18	Female	18	96	6
4	Female	16	95	6
38	Female	19	95	6
86	Female	19	95	6

30	Female	18	94	6
44	Female	16	94	4
62	Female	19	94	6
26	Female	19	93	6

All the top ten performers studied for at least 4 hours, with 9 studying for 6 hours.

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

Correlation = 0.764358

There is a positive correlation between more study hours and higher exam scores

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

   Generally, there is a positive correlation between studying more and getting a higher test score. The female students tended to study more than their male counterparts, and their scores reflected their effort. Surprisingly, there was little to no correlation between age and exam score.

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.