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	85.01111	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

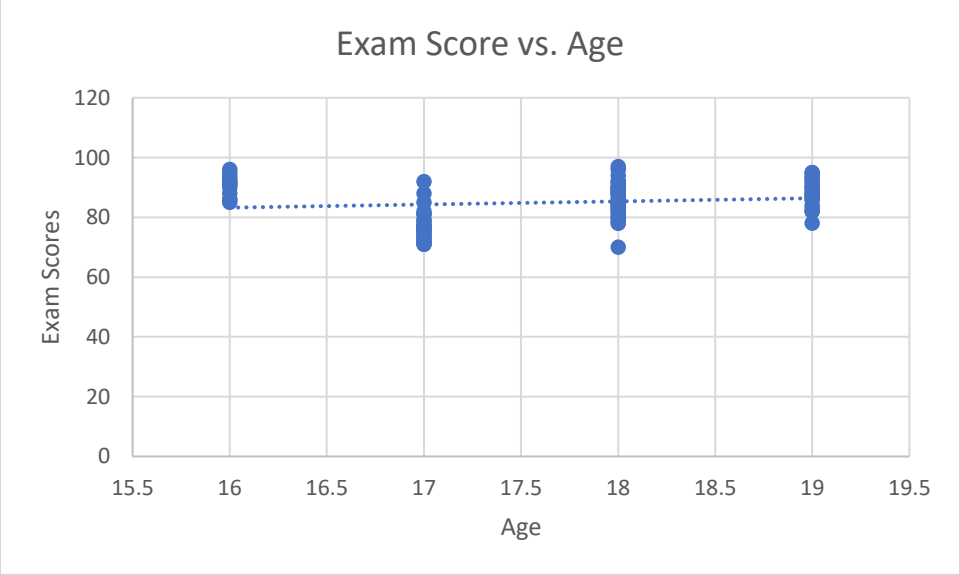
Minimum	70	Minimum	2
Maximum	97	Maximum	6
Sum	7651	Sum	402
Count	90	Count	90
Confidence Level(95.0%)	1.444443	Confidence Level(95.0%)	0.239526

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**

Female study hours are on average an hour more than males meaning that the average exam scores are higher than males.

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



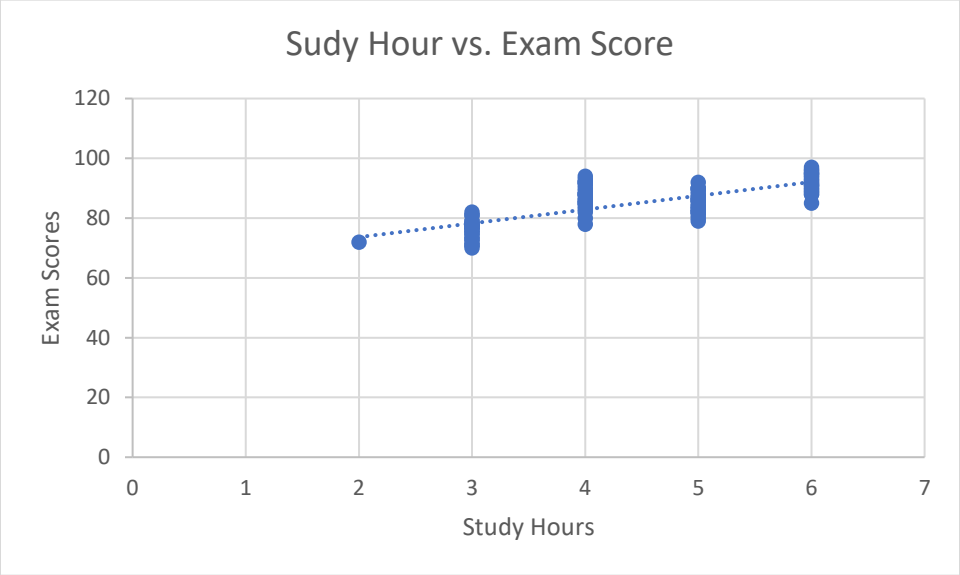
There is a 0.15 correlation between age and exam scores. This means that there is a weak correlation between age and average exam, meaning that age does not affect score exam.

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**

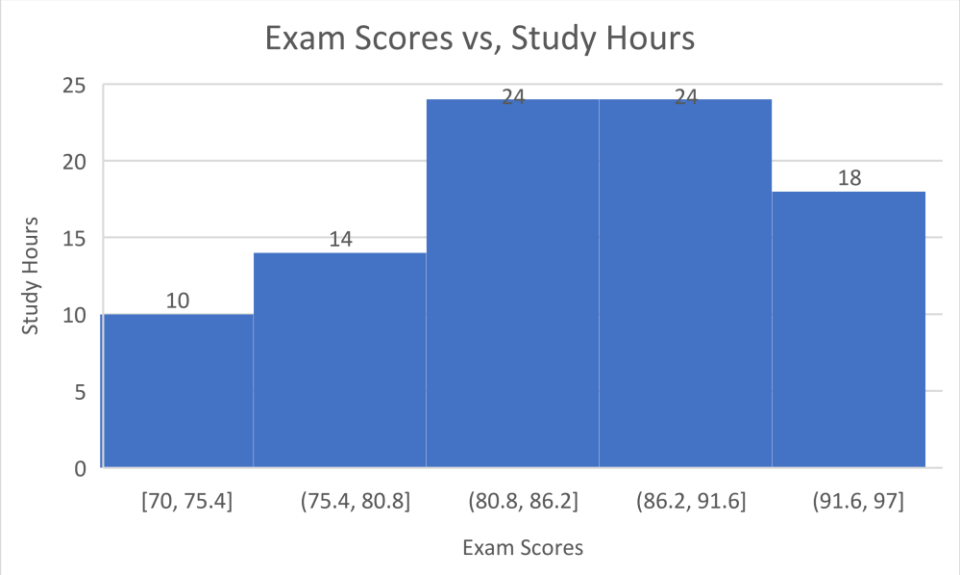
Math and Science are the strongest subject and English is the weakest subject.

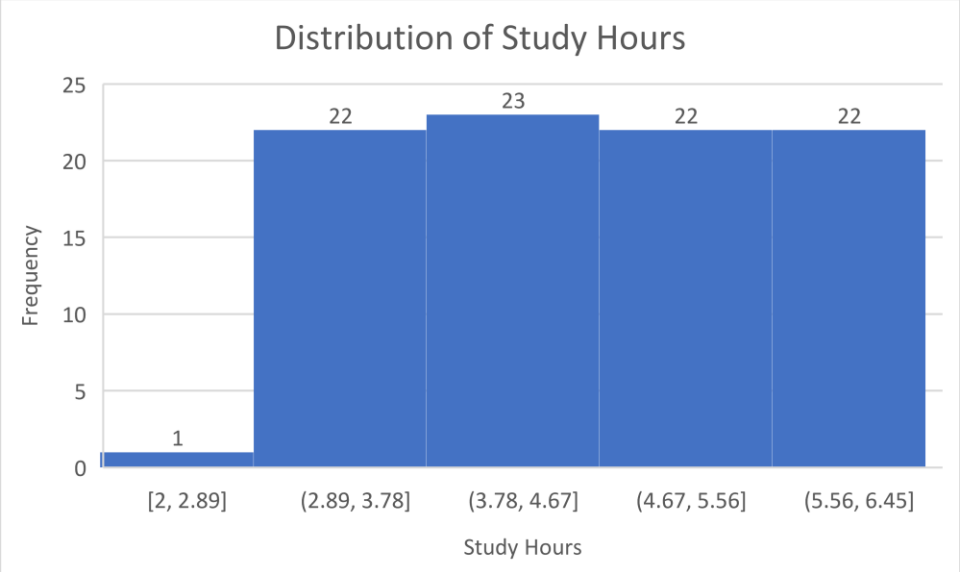
E. **Study Hours vs. Exam Score:** Create a scatter plot to visualize the relationship between study hours and exam scores.



The relationship between study hours and exam scores seems to be highly correlated, meaning that there is a positive correlation of 0.76 between study hours and exam scores. This means that the more you study the better the scores you get.

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





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

Student ID	Gender	Age	Subject	Exam Score	Study Hours
90	Female	18	Science	97	6
8	Female	16	Science	96	6
18	Female	18	Science	96	6
4	Female	16	Math	95	6
38	Female	19	Math	95	6
86	Female	19	Math	95	6
30	Female	18	Science	94	6

44	Female	16	Math	94	4
62	Female	19	Math	94	6
26	Female	19	Math	93	6
52	Female	16	English	93	4
78	Female	19	Science	93	6

The top students are all females of averaging ages between 16 to 19 years. There is a weak to no correlation of 0.38 for study hours and grade and there is a weak negative correlation of -0.11 between age and grade. Most of them study 6 hours except for 2 of them.

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

   The relationship between study hours and exam scores seems to be highly correlated, meaning that there is a positive correlation of 0.76 between study hours and exam scores. This means that the more you study the better the scores you get.

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.